

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2002, 23:25:50 ; Search time 2520.32 Seconds
(without alignments)
5965.754 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attcagcatatagatgt.....gagagggttaattcattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	199.6	17.9	725	12	AQ078448 CIT-HSP-2
2	188.2	16.9	690	12	AZ840584 2M0138N11
3	184.6	16.6	777	12	AZ741291 RPCI-24-9
4	159.6	14.3	808	12	BH039878 RPCI-24-2
5	153.2	13.8	729	12	BH083291 RPCI-24-3
6	141.8	12.7	617	12	AZ795382 2M0049N06
7	141	12.7	545	12	AZ748309 RPCI-24-6
8	135.2	12.1	635	12	BH062438 RPCI-24-3
9	132.2	11.9	574	9	A1806860 wf36807.x
10	129.4	11.6	615	12	AZ553624 RPCI-23-2
11	128.8	11.6	610	12	AZ222872 RPCI-23-4
12	128.4	11.5	467	12	AQ117475 HS_2187.B
13	127.4	11.4	563	12	AZ810472 2M0076E06
14	127	11.4	638	12	AZ553591 RPCI-23-2
15	126.8	11.4	740	12	BH267496 CH230-82L
16	119	10.7	853	12	AZ265296 RPCI-23-1
17	118.2	10.6	493	12	AZ759046 1M0551P17

18	116.4	10.4	447	9	AA012849
19	116.4	10.4	768	12	BH060668
20	114.2	10.3	597	12	AZ422002
21	113.4	10.2	715	12	AG013852 Homo sapi
22	112	10.1	643	12	BH048022 RPCI-24-2
23	112	10.1	698	12	BH039875
24	111.8	10.0	397	9	AA021085
25	111.2	10.0	424	10	H86939
26	110.2	9.9	586	12	AZ093307 RPCI-23-4
27	108.6	9.7	717	12	AG013832 Homo sapi
28	108	9.7	682	12	AG130179 Pan trogl
29	105	9.4	511	12	B67083 CIT-HSP-201
30	104	9.3	495	12	AQ223006 HS_2017.A
31	100	9.0	243	12	AZ737781 RPCI-24-8
32	98.4	8.8	716	12	AG013834 Homo sapi
33	97.8	8.7	715	12	AZ994117 2M0279M22
34	96.4	8.7	696	10	BE968588 601649752
35	95.6	8.6	725	12	AG013831 Homo sapi
36	94.6	8.5	382	9	AA442630 zV59F10.r
37	92.8	8.3	677	12	BH267500 CH230-82L
38	91	8.2	580	9	AW973537 EST385637
39	90.4	8.1	481	12	B92737 CIT-HSP-217
40	87.8	7.9	406	12	AZ767149 1M0566H11
41	86.8	7.8	597	12	AZ896363 RPCI-24-2
42	86.4	7.8	593	12	AZ754210 RPCI-24-6
43	84.6	7.6	207	12	B59457 CIT-HSP-201
44	84	7.5	658	12	AQ666848 HS_2105.A
45	83	7.5	675	12	AG000745 Homo sapi

ALIGNMENTS

RESULT 1

AQ078448 725 bp DNA linear GSS 20-AUG-1998
CIT-HSP-2358C21.TPB CIT-HSP Homo sapiens genomic clone 2358C21, DNA sequence.
AQ078448
AQ078448.1 GI:3439632
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 725)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: CIT-HSP-2358C21.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdaams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

JOURNAL COMMENT

FEATURES source

1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2358C21"
/sex="Male"
/cell_type="Sperm"

/note="vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 190 a 184 c 120 g 231 t

Query Match 17.9%; Score 199.6; DB 12; Length 725; Best Local Similarity 63.0%; Pred. No. 8.4e-38; Matches 345; Conservative 0; Mismatches 194; Indels 9; Gaps 2;

QY 15 agatgttgaaattggtattatgagacaatgagcagaaattatgctattctcattagatc 74
Db 76 AAAAGATGATGATGGTAAACAAATTTGAACACATGAACCAACCAATGATAACATGGGATC 135
QY 75 tcttgctt-----tctccacagatgatactcttgcttgaatttcttcccttcaagatgata 128
Db 136 TATTGTTTCTTTTATCCAGATATAGAAAGATGTTTCTCTATGCCATTAAAGATGAAC 195
QY 129 aactctcgtgtttattacatacaaatctcttttttttttttttttttttttttttttttttt 188
Db 196 TATTTTCTACTTATGCTATCGTTAAATAATGCCTTTGTTTTCAGCTGGCATTTGGATTC 255
QY 189 cagcaacaacacacctctctcttt 248
Db 256 CAGGCAACACCTTCTCTCTTTTCTTCCACATCTTTAGCTTCTTCTGAAATCGCAGGCTA 315
QY 249 agtcattgacatgataattagtagtaacctgtctctctccacacatactgctctctcactc 308
Db 316 AACCCCGTGACTTACTCACCTGTCACTGGCCCTCAFTCAATTCAGATGCTCCTCACTG 375
QY 309 agcaaatatgggtccttagactcttttttttttttttttttttttttttttttttttttttt 368
Db 376 CAGTGGATTTTGGCCCTAGACATATTTGAATCACTGCAATTTTGGGAATGACTTTAAGT 435
QY 369 ataaggtcattgtcttt 428
Db 436 GTAAGGACATTTTACACAAACAGGGCAATGAAGGCCCTTCACTGCAACACCTGCC 495
QY 429 tctcgtgagtgctccagc---catcatcagcccccagcatctctctctcttgccaaagctca 495
Db 496 TTCTGAACATGCTCCAGGCCATCATCATCAGCCCGCAGCACCTTCTGTTGGCAAGGTTA 555
QY 486 acatctctcgaagtcacatcttaggattctctctctctctctctctctctctctctctctct 545
Db 556 ACATAAATCCAAATATACATTTTGCATGTTTCTCTTTTGGGTGCTGCTCAATTTGT 615
QY 546 tcattggt 553
Db 616 CCTTCACT 623

RESULT 2
AZ840584
LOCUS 2M0138N11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0138N11 F, DNA sequence.
ACCESSION AZ840584
VERSION AZ840584.1 GI:13010492
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 690)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0138 row: N column: 11
Seq primer: CGTGTAAACAGCAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 690.
Location/Qualifiers
1. .690
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (G14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES
source

BASE COUNT 174 a 184 c 118 g 214 t
ORIGIN
Query Match 16.9%; Score 188.2; DB 12; Length 690;
Best Local Similarity 60.6%; Pred. No. 4.8e-35;
Matches 396; Conservative 0; Mismatches 248; Indels 9; Gaps 5;
QY 149 atacaattctttttatcacaaagcagcttgggaatttcagcaaacacattcttctt 208
Db 33 ATAGGAAACATTTTCTGAAATGGCATTTGGGATCTCAGGCAACAGCTTCCCTCTT 92
QY 209 cttttccacattctcacttctgttttttcagtcacaggtcttaagtcattgacatgataatt 268
Db 93 CTGTGTCACATCTTAAGTTTCATTCGTGGCAGCAGGCCAGACTACTGACCTGCCATT 152
QY 269 agtcacctgtctctcatccacatactgctctcttctcactcaggaataattggttcccta 328
Db 153 GGTCTCTTGTCCCTTAATCCACCTACTGCTACT---GGTCGACGATTCATAGGCACA 209
QY 329 gactcttgggttcacagaatactcaggtatgcttaggtataaggtcattgctctttta 388
Db 210 GACATTTTATCTCTCGGAGGGCTGGGATGACATCATATGTAAATTCCTGTGTACCTG 269
QY 389 acaaggtgatgaggggcctctccatctgcacccccctgcctcctgagtggtgtccaggcc 448
Db 270 TACAGAGTTTGAAGGGTTCTCCCTTTCACCCAGCAGCATGTTGAGTATCTCTCCAGGCC 329
QY 449 atca-----tcagccccagcatcttctccttggcaaaagctcaaacatctcttgcagtcac 505
Db 330 ATCATTTCTTAGTCCCAAGCTCTCTGTTTAGCAAAAGTTCAAGCATATATCTCTCATCAC 389
QY 506 atcttaggattcttctcttct 564

Db 390 ATATCAAGTGCATCTTTTCTCAGTGTCTCTATATGTTAATGGCAGTCAACTCTTG 449

Qy 565 ctgtacacgtgggtacacccagtgaaacgggagctctctctgtt-tgtcatcacagcac 623

Db 450 GTATCCATCATTTGACACCCCAATTTGACCAATGAATGACTTTATTTATGTTACTCAGTCC 509

Qy 624 ttttctcttttggccatgagctacacccacagaggaactgtttttcacactaaatgactttg 683

Db 510 TGCTCTATTCTACCCCTTGAGTGTACTCATGCAAAAGCATATATTCTACACTGCTAGCCATC 569

Qy 684 agggatgtcacctttttaggtttctctctcaagagagctacatgggtgattatttta 743

Db 570 AAGCAATCTTTCTATTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 629

Qy 744 tacagacaataaagggtctctctcagtccttcacgcgcgcagcgtgtcccccga 796

Db 630 AGCATGC-ACAGGAACAGACACAGCATCTTCTATGGAACCAACCTGTCCCCCA 681

RESULT 3

AZ741291/c A741291 777 bp DNA linear GSS 25-JAN-2001

LOCUS RPCI-24-90L5.TJ RPCI-24 Mus musculus genomic clone RPCI-24-90L5,

DEFINITION DNA sequence.

ACCESSION A741291

VERSION A741291.1 GI:12516500

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 777)

AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: RPCI-24-90L5.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 90 row: L column: 5

Seq primer: SP6

Class: BAC ends.

FEATURES

source

1. .777

Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-90L5"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/notes="vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 260 a 141 c 154 g 222 t

ORIGIN

Query Match 16.68; Score 184.6; DB 12; Length 777;

Best Local Similarity 60.94; Pred. No. 3.7e-34;

Matches 388; Conservative 0; Mismatches 239; Indels 10; Gaps 5;

Qy 385 tttaacaagaatgatgaggggctctccatctgcacccctgcctcctgagtggtctcca 444

Db 774 TTTAAACAGGGCAATGAGAGGTCTCTATCAGCATCACCTGCCTCTAGAGTGTGCCAA 715

Qy 445 ggcacatc---atcagcccccagcatctctctctctggcaaaagctcaaacatcctcttgaag 501

Db 714 GCGTGTGACGATAGTCCCACTCCGCTTTGTGGCAACAATTTAAACATAAAACAAAA 655

Qy 502 tcacatcttaggatctctctctctctctcaacatgcttcatttggtgtaactctt 561

Db 654 ATACATGATCTATTCTTCTCTATTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 595

Qy 562 ctgct---gtacactgcgcgtacaccccccagtgaaacgggcccagttctctctgttctacata 619

Db 594 GATCTTCTATACTGGTGTCTATATCAACGCTGAGTGAGACCAACGAGTGAAGGTCACATA 535

Qy 620 gcactgtctcttttggccatgagctacacccacagagagactgtttttcacactaaatgac 679

Db 534 ATACTGCTACTCTTCCCTATGAACAACATCATCAGGGGACTAATTTCAACAATCATTAAC 475

Qy 680 tttagggatgtcaccttttaggggttcattggtctctctcaagagggctacatggtgattat 739

Db 474 CTTAAGAGATGTATTTTGTAGGAGTCATGCTGAGCACAAGTACATACATGTTGATTAT 415

Qy 740 ttatcacagacaataaagggtctctcagtcgcttcacgcgcgcagcgcctgtccccgagtc 799

Db 414 CTTGTTTCAGAC-ATCAGAGGCAATGCAAGCATCTTCTATAGCGACAGCCACTT-GAGAGCC 357

Qy 800 tcaccagtgaagaagagctccacggctctctctctctctctctctctctctctctctctca 856

Db 356 TCCCTCTGAGAAAGAGCCACCCAGACCATCTGCTGCTGGTGTCTTCTATGTTGGTCAAG 297

Qy 857 tactgggtggacttttaagct 916

Db 296 TACTGGGTGGACTTTTATGTCATCTAGAACAGTCTTTTATGATGATGATGATGATGATG 237

Qy 917 ctagtgtggtccagggtattgtgtgcccataagctatgcgcgcgcgcgcgcgcgcgcgcgc 976

Db 236 ATGCTTACTGTTCAGAAAGTTTGTGTTGAATGCCCTATCCCAATTTACTCTCTTAATACAA 177

Qy 977 attatgctgataaacaataattcaagactctgcaaa 1013

Db 176 ATTAGTCTGTATAATCGAATCTCATGATTTCTATAAAA 140

RESULT 4

BH039878

LOCUS RPCI-24-273F22.TV RPCI-24 Mus musculus genomic clone RPCI-24-273F22

DEFINITION 'BH039878 , DNA sequence.

ACCESSION BH039878

VERSION BH039878.1 GI:14818265

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 808)

AUTHORS Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Akinret,B., Levins,M., Russell,D., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: RPCI-24-273F22.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 273 row: F column: 22

Seq primer: T7
Class: BAC ends.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

1. .808
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-273F22"
/clone_lib="RPCI-24"
/sex="Male"

```

```

/sex= male
/cell_type="Spleen/Brain"
/notes=vector: pTRAPAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRAPAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA.

```

DNA.			
BASE COUNT	234 a	172 c	139 g
ORIGIN			
			263 t

Query Match 14.3%; Score 159.6; DB 12; Length 808;
Best Local Similarity 55.9%; Pred. No. 4.3e-28;
Matches 389; Conservative 0; Mismatches 294; Indels 13;

[illegible]

Qy 388 aaacaaggatgatgaggggacctctccatctgcaacccctgcctcctgagtgctccaggc 447
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 99 AAACAGATGATGAGAGGCCCTCTCTATCTGCATCACCTGCCTCTGAGTGTATCCAGGC 158

Qy 448 catc--atcagccccagcatcttctcccttgcaaaagctcaaacatccttctgcagtca 504
||| ||||| ||||| | | ||||| | | | | | | | | | | | |
Db 159 TATCCCAATCAGTCCCAGTAGCTCTTTGTCGAAAAATTAAAGATAAAGTAACAAAACA 218

Qy 505 catcttaggatttctcctttttctcatgggttcctcaacatgttcaattgggtgtaattctctg 564

Dβ 219 GATCATCTACGCTTTTTCATCTCTCATCTCTTCATTTCAATTTGT --CATTCAGTAGTAGGCGG 276

Qy 565 ctgtacactcgggtaccccccaagtgaacggggccagttcttctgttctatcacgacct 624

Dp 277 ATCTTCCATGTTGTGTCATTTTACCAGATGTGAAAACTACACATACACGCTCATTAATCTT 336

[illegible]

Qy 685 gggatgtcaccttttatagggttcattggtcctctcaagggtcacatgggtgattatttat 744

Qy 745 acagacataagaggctatctcagtgccttcacgcagcagcctgtccccgagtcacc 804

45/ GCATACATCAGCAATGCAAGGATTTTCATAGGATCAGCCA-----CTGAGAGGCACTCC 511
 UU
 Qy 805 agtgaaagagcctcccagggtatcttactgctggtgagt---tttgctcttcacatactg 861

DG 512 TGAGAAAAGGGCCACCCAGACCATTCTTCTTTGTGGTTATGTA CTG

QY 862 gttggactttagcttcctcattttcaggaggtgacatggataaatgattctctgctagt 921

Accession	Sequence	Position
Db	572 GTTGGACATCATCATCTCAATCACCTCAATCTGTTTGGATGATGACTCAGTCATTCT	631
Qy	922 gtggctcagggtattgtggccaatagctatgccgcgaattagtcctttgatgctcaattta	981

Db 632 GACCATTCAGCAGTTTGATGATGCTATGCCATATCCCACAATTGCTCCTTTGGTACAGCTCAG 691

Qy	982	tgctgataaccaaattattcaagactctcgaaatgtt	1017
Db	692	TTCGTATAGAGAAATATCCGTATACTAAAAAATT	727

RESULT 5

BH083291
LOCUS
DEFINITION
729 bp DNA linear
BH083291
RPCI-24-309J17.TV
Mus musculus genomic clone
RPCI-24-309J17
, DNA sequence.

ACCESSION BH083291
VERSION BH083291.1 GI:14902888
KEYWORDS GSS.

SOURCE
ORGANISM

ORGANISM

MUS MUSCULUS

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus

1 (bases 1 to 729)

REFERENCE

AUTHORS

Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,

TITLE
JOURNAL
COMMENT

CONTACT: Shaifang Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhae@tigr.org
 Fax: 301 838 0208
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.htm.
 Plate: 309 row: J column: 17

Seq primer: T/
Class: BAC ends

URES
SOURCE

```

Location/Qualifiers
1. .729
/organism="Mus musc
/strain="C57BL/6J"
/db_xref="taxon:100
/clone="RPCI-24-309
/clone_lib="RPCI-24
/sex="Male"

```

/cell_type="Spleen/Brain"
 /note="Vector: pTABAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

DNA.	
BASE COUNT	ORIGIN
170 a	181 c
138 g	240 t

Query Match 13.8%; Score 153.2; DB 12; Length 729;
Best Local Similarity 58.2%; Pred. No. 1.5e-26;
Matches 384; Conservative 0; Mismatches 263; Indels 13;

Qy 78 Tgcttttctccacagatatacctcttggtttaatttccttctaagtatgatcaaacctctcg 137
||| ||| | | | | | |||| | | | | |
Db 74 TGTTCTTTTTCTCCGTTTCACATGTTTCACATAAATGAATAAGAACAGCAGACTACATA 133

Qy 138 gtttattaccatacaaaatcttcttttatccacaagcagcttggaatttcagaaaac 197
| | | | | | | | | | | | | | | |
Dh 134 CTAAATTCCTACATTGCGCCAGC-ACACTTCTTTCTCTCACTTCTTGCCCATGCACAGCACACA 192

[illegible]

DEFINITION RPCI-24-63G18.TV RPCI-24 Mus musculus genomic clone RPCI-24-63G18,
DNA sequence.
ACCESSION A2748309
VERSION A2748309.1 GI:12533043
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 545)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-63G18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pjejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 63 row: G column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-63G18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 160 a 116 c 112 g 157 t
ORIGIN

Query Match 12.7%; Score 141; DB 12; Length 545;
Best Local Similarity 60.9%; Pred. No. 1.2e-23;
Matches 333; Conservative 0; Mismatches 205; Indels 9; Gaps 6;
QY 301 cttcactcagcaaatattggtcttagactcttttgggtccacagaataactcaggatga 360
DB 1 CCTCATTGGAGGAATGTCATGCTGCGACATATTTAGTCACATGAATGTCAGAAATGA 60
QY 361 tcttaggtataaagtgcatgtctttttaacaaggtgatgagggccctccatctgcac 420
DB 61 CATCAATGTATAGCAACTTATACACAAACAGGGTGATGAGAGGCCCTCTCTATCAGCAT 120
QY 421 cccctgctcctgagtggtgtccaggc---catcatcagccccagcatcttctccttggc 477
DB 121 CACCTGCCTCCTGAGTGTGATCCAGGCTGTGCACAAATCAGTCGCCAGTACCTTTATGGTGCA 180
QY 478 aagctcaa-acatcctcttgcaagtcacatcttagattcttctcttcttcctatgggtcc 536
DB 181 AAATTTAAACATAAATCAAGAAACACATGGTCAATGCTCTTTCTTTATATTTGGTCT 240
QY 537 tcaacatgt--tcattggtgtatcttctgtgtacactcgcgtaccctccagtg-aaac 593
DB 241 TTCAATTTTCTCTTAGTAGTACCTCGAUCTTCTATACTGCTGGTTTACCAATGTGAGT 300

QY 594 ggggcccagttctctgtttgtccatcacagcactgttctcttttggccatgagctacacccac 653
DB 301 GAGACCAAGCAGATGAAGATCACTAAGTCTGCTCAATCTTGCCAAATGAACATACATCATC 360
QY 654 aggagactgtttttccacactaatgacttttggggatgtccacctttataggggttcagtgc 713
DB 361 AGGGGAATGTTGTTAAACAGTGCACAACCGTCGAGAGATGTTTCTTGTGGGAGTTATGCTG 420
QY 714 ctctcaagaggtcacatgtgtgattttttatcacagacaataagaggtctatctcagtgct 773
DB 421 ATCACAAGTCATACATGGTGATATCTTGTTCAGAC-ATCAGAGGCAATGCAAGCATCT 479
QY 774 tcacgcagccagcgtctgcccgcaggtctccacagtgaaagagcctccacgtctatcttac 833
DB 480 TCATAGCATCA-ACAACCTGAGAACCTCCCTGAGAAAGGGCCACACAGACCATCTTGC 538
QY 834 tgctggt 840
DB 539 TGCTAGT 545

RESULT 8
LOCUS BH062438 635 bp DNA linear GSS 18-JUL-2001
DEFINITION RPCI-24-327B18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-327B18
ACCESSION BH062438
VERSION BH062438.1 GI:14875948
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 635)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pjejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 327 row: B column: 18
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..635
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-327B18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 174 a 130 c 125 g 206 t
ORIGIN

Query Match 12.1%; Score 135.2; DB 12; Length 635;

```
Best Local Similarity 61.58; Pred. No. 3.2e-22;
Matches 252; Conservative 0; Mismatches 153; Indels 5; Gaps 2;

Qy 607 tgtttgtacacagcactgttccttttggccatgagctacacccacaggagactgtttt 666
Db 3 TGAAGGTGATTAAATACGTCTCACTCTCCCACTGAACATACATCAGGGGACTGATTT 62
Qy 667 tcacactaatgactttgaggatgacacctttataggtttcataggtctctcaagagct 726
Db 63 TAACCATGACAACTTAAGAGATGCTGTTCTGTAGGAGTCATGCTGACTACGAGACAT 122
Qy 727 acatgggtattttttacacagacaataagaggctatctcagtcgcttcacagcagc 786
Db 123 ACATGGTGATTATCTGTTTCAGACATCAGAGGCAATGCTTTTTTCTCATAGCATCAA-- 180
Qy 787 ctgtcccgagctctcacagtgaaagagcctcccgagctatctactgtcgtgagtg-- 844
Db 181 CCACCTGAGAGCCCTCCCTTGAGAAAGGGCCACTCATACCATCTTGCTGCTGGTGT 240
Qy 845 -tttgtctcacactactgggtgactttacgttctctctatcttctcagggaggtgacatgat 903
Db 241 CTTTGTGTCATGTAAGTGGGAGGACTTCATCATCTCATFCCACCTCAGTCCCTGTATGGAT 300
Qy 904 aaatgattctctgctagtggtgctccagggttattgtggccaatagctatgcccgaattag 963
Db 301 GTATGACCCAGTCATCTGACTGTTTCAGAGGTTTGTGATGAATGCCATCTACAAATTAC 360
Qy 964 tcccttgatgctaattatgctgataccaaaataattcaagactctgcaaa 1013
Db 361 TCCATTGTGTACAAATCAAGTCTGTGTAAGCGAATAATCATGATGCTGCTAAAAA 410

RESULT 9
A1806860/c
LOCUS
DEFINITION
w36807.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2357653 3' similar to TR:Q62855 Q62855 PHEROMONE RECEPTOR
VN6. ; mRNA sequence.
ACCESSION
A1806860
VERSION
A1806860.1 GI:5393426
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 574)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 659 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 447.
Location/Qualifiers
1. 574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2357653"
/lab_host="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site:1; Not 1; Site:2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHR, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
```

```
I.M.A.G.E. clones 297480-302087, 682632-687239,
728408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 179 a 108 c 138 g 149 t
ORIGIN

Query Match 11.98; Score 132.2; DB 9; Length 574;
Best Local Similarity 60.48; Pred. No. 1.6e-21;
Matches 306; Conservative 0; Mismatches 193; Indels 8; Gaps 5;

Qy 514 attcttcttttctcaggggtcccaacatgttctcatgggt-gtaattcttctgctgtacac 572
Db 566 ATGTTTCTTCTTCTATGGTTCCCTCACTTGTCTTTCAGTAGTAACAGGATCTTCTACAC 507
Qy 573 tgcggctcaccccgagtg-aaacggggccagctctctctgtttgttcatacagacacttcttc 631
Db 506 TGTAGCTGTTTCCAGTGTGAGCCAGACAAATGTACTGAATATAGGTAACTACTGCTCAGT 447
Qy 632 ttggcccatcagctacacccacagagagactgttttccacactaatgactttgagggatgt 691
Db 446 TTTCTCATGAGCTTCATCTCAGGGGCCACGTTTCTTATTCTGACAACTTAGGGATGC 387
Qy 692 cacctttatcaggttctcaggtccctctcaagaggctacatggtgattattttatacagaca 751
Db 386 TTTCTGAGTAACTACTGCTGTTCTTAATTGCATACATGCTGACTCTCTTATTTCAGGC- 326
Qy 752 ataagaggctatctcagtgcttcacgcagcagcctgtccccc--gagctccacagtgga 809
Db 327 CTCAGAGGTGATCCCAACACCTTCACAGCAATGGTCTTCCCAAGGAGCTTCCCTCTGAGA 268
Qy 810 aaagagctcccgaggtatcttactgctggtgagttt---tgtttcacacatactgggtgg 866
Db 267 AACGGGCCATCCAGACCATACTGCTGTGGTGAAGTTCTCTGTGGTCTCTGTGGGTGG 208
Qy 867 actttagcttctctatttcaggaggtgtgacatgataaaatgattctctctgctagtgtgc 926
Db 207 ATTTTCATCATCTCTCTTCTCCTCAGCAATGTTGTGGCATATGCGCCCAATCAACGCAAG 148
Qy 927 tccagggtattgtggccaatagctatgcccgaatagctcctttgatgctaattatgtctg 986
Db 147 TTCAGAGTCTTGTAGTCAATGCTATGCCGTGATCAGTCCCTCTGTGTCTCTCTCTCAG 88
Qy 987 ataaccataattcaagactctgcaaa 1013
Db 87 ATAAAGAAGAACAACCAATATTCTGAAAA 61

RESULT 10
A2553624
LOCUS
DEFINITION
RPCI-23-209L1.TV RPCI-23 Mus musculus genomic clone RPCI-23-209L1,
DNA sequence.
ACCESSION
A2553624
VERSION
A2553624.1 GI:11233124
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 615)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-209L1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
```

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 209 row: L column: 1
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..615
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-209L1"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 178 a 128 c 107 g 202 t
ORIGIN

Query Match 11.6%; Score 129.4; DB 12; Length 615;
Best Local Similarity 58.3%; Pred. No. 8e-21;
Matches 284; Conservative 0; Mismatches 196; Indels 7; Gaps 3;

Qy 99 ttctgttaatttctcttaagatgacaaactctctggtttattaccatacaaatct 158
Db 80 TATGACTCAACAAGCTCTTAAGATAAATTCACCTCTCACTATGTTCTCATGGAGAAT 139
Qy 159 tct-tttatccaaagccagctgttgggaatttcagcaaacaccatctcttctttccac 217
Db 140 GCTCTTTATATCCAAAGCTGGGTAGGAGTCTAGTCTTAATATGTCCTCTTTCTCTAT 199
Qy 218 atcttcaccttgtttcagtcacaggtctcaagtcacattgacatgaataattagtcacctg 277
Db 200 ATTTTCA---TAATCCTTAGGCGACAGACCTTAAGCCCAATGACCTTAATCTCCTGTCAACAG 256
Qy 278 tctctatccacatactgctgctcttcaactcaggcaataattggtctcttagacttcttt 337
Db 257 ACTTTTCAATACATAATGCTGCTTCTCAGTCGAGGATATTTTGCATACAGATATATTT 316
Qy 338 ggttcacagaatactcaggatgatcttaggtataaagtgcatgttctttttaacaagggtg 397
Db 317 GAGTCAATGAATATTGAGATGACTTCAAAATGCAAAACAACTTTTACATATGCAGGGTA 376
Qy 398 atgaggggctctccatctgcacccctgctctgagtgagtgctccaggc---catcatc 454
Db 377 ATGAGAGGCTCTCTATCTGCACCACTGCTCTGAGTGTGTTCAGGCTGTCCACATC 436
Qy 455 agccccagcatctctcttggcaagctcaaacatctctctgcaagtcacatcttagga 514
Db 437 AGTCCCAATACCTTCGCTGTGGCAAAATTTAAACATAAATAAATAACATATCAAA 496
Qy 515 ttcttctttctcatgggtctcaacatgttcatgttggtgtaattcttctgtgacactg 574
Db 497 TGCCTTCTCTATATTTGGGGCTTTAAATTTGCTCTCAGTAGTAGTAACCTGTATCTTCATG 556
Qy 575 cggctac 581
Db 557 TTGGTGC 563

RESULT 11
AZ222872/c

LOCUS AZ222872 610 bp DNA linear GSS 14-JUN-2000
DEFINITION RPCI-23-46M10.TJ RPCI-23 Mus musculus genomic clone RPCI-23-46M10,
DNA sequence.
ACCESSION AZ222872
VERSION A2222872.1 GI:8530921
KEYWORDS GSS.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 610)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-46M10.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 46 row: M column: 10
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..610
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-46M10"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 194 a 114 c 124 g 178 t
ORIGIN

Query Match 11.6%; Score 128.8; DB 12; Length 610;
Best Local Similarity 59.3%; Pred. No. 1.1e-20;
Matches 273; Conservative 0; Mismatches 182; Indels 5; Gaps 3;

Qy 574 gcggctacccccagtgaaacgggcccagctctctgttgcatacagcactgttcccttt 633
Db 590 GTGGTTCCCAATGTGAGTGAGCAACCAACAGATGCAGGTCACCTAAATCTGCTCACTCT 531
Qy 634 tgccatgagctacacccacagagactgtttttcacactaagactttgaggatgca 693
Db 530 TACCCATGAACATATCATCAGAGGATTTGATTTTACAATATCAACCTCCAGAGATGTT 471
Qy 694 cctttatagggttcactggtctctcaagagctcacatgggtgattattttatacagacaat 753
Db 470 TTGTGTAGGAGTCATGCTGACCACAAAGTGTATACATGGTCATTTACATTTTAT 412
Qy 754 aagaggctatctcagtccttcacgcagccagcagctgtccccgcagctctccacagtgaaaag 813
Db 411 CACAGGAGTGCAGAGTATCTTTATAGCATCAGCCACT-TGAGAGAGTCCCTTGAGAAAG 353

```

Qy 814 agcctccaggctacttactgctggtagt---tttgcttcacatactggtgagctt 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GCCCACCAGACCATCGTCTGCTGGTGTCTTTCTTTGTTGTCATGACTGGGAGACTT 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 871 tacgtctcatttcaggagdgctgacatgataaatactctgctagtggtgctcca 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 CATCATCTCAATTCACCTCAGATATGATGATGATGATGATGATGATGATGATGATG 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 931 ggttatgtggccaatagctatgcgcgaattagtcctttgatgctaaatttatgctgataa 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GAAATTTATGATGATGCTTATCCACCAATTTACCCCTTTGGTGAACAATTCAGTTCTGTATA 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 991 ccaaatattcaagctctgcgaatggttatggttttaaatat 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 CAGAAATAATATTGCTGAACCACTGCAATCAAAACAT 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AQ117475
LOCUS
DEFINITION
  HS_2187_B2_D03_MR CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate-2187 Col-6 Row-H, DNA sequence.
ACCESSION
  AQ117475
VERSION
  AQ117475.1 GI:3495266
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 467)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Sequence Tagged Connector
  Plate: 2187 row: H column: 6
  Class: BAC ends
  High quality sequence stop: 467.
  Location/Qualifiers
    1..467
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate-2187 Col-6 Row-H"
      /clone_lib="CIT Approved Human Genomic Sperm Library D"
      /sex="male"
      /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
      E-Coli DH10B"
BASE COUNT
  121 a 114 c 82 g 148 t 2 others
ORIGIN
  source
    1..467
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUC2M0076E06"
      /clone_lib="Mouse 10kb plasmid UUC1M library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
      /note="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA

```

```

Qy 498 caagtcacatcttaggattcttctcttctcattgggtcctcaacatgttcattggt-gta 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 CAAATACGATATCCTCGCTGCTTATTCGTTTTTGTAGCAATTTGCTTCAGTAGTG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 557 atctctgctgtacactgcggtctaccctccagtgaaacggggccagctctctgtttgtcat 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 ACATGATAATCTACACTGTAGGTTATTCCAATGACCCAG----ATAATTTTGAATATCAG 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 617 acagcaactgttctctttttgccccatgagctacacccacacaggagagctgtttttcacacactaa 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 CAAATATTTGCAATATTTTCCCAATGAATGCTCTCATCAGGACGCTATTTCTTATGCTGTGTC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 677 gactttgaggtatgcacctttatagggttcatggtctctcctcaagaggctacatggtgat 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 ATTATCCAGAGATGCCCTTCTTCATAGGAATCACCGCTGCTCTCAAGTGTATACATGGTTCAT 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 737 tattttacagacaataagaggctatctcagtccttcacgcagccagcctgtctcccca 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TCTTTTGTCCAGGC-ATCAGAGGCACCTCCAGCAGCTTTTCACAGCAGCAGCCTTATATTAA 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 797 g--tctcaccagtaaaagagcctcccgaggtctatcttactgctgtga 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GGACTTCTCTAGTGAATAATGGCCACCACCAAGACCATCTCTGATGCTGTGTA 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AZ810472
LOCUS
DEFINITION
  2M0076E06F Mouse 10kb plasmid UUC1M library Mus musculus genomic
  clone UUC2M0076E06 F, DNA sequence.
ACCESSION
  AZ810472
VERSION
  AZ810472.1 GI:12977756
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 563)
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0076 row: E column: 06
  Seq primer: CGTTGTAACAGCAGGCCAGT
  Class: plasmid ends
  High quality sequence stop: 563.
  Location/Qualifiers
    1..563
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUC2M0076E06"
      /clone_lib="Mouse 10kb plasmid UUC1M library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
      /note="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 01:55:45 ; Search time 376.86 Seconds
(Without alignments)
5075.207 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attccagatcatagagatgt.....gagaggttaattcattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: * 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231.4	20.8	1331	18 AAT69546	Rat pheromone rece
2	226.4	20.3	1538	18 AAT69549	Rat pheromone rece
3	221.8	19.9	1386	18 AAT69545	Rat pheromone rece
4	215	19.3	584	22 ABA60235	Human foetal liver
5	215	19.3	584	22 ABA28536	Probe #7002 for ge
6	215	19.3	584	22 AAK08511	Human brain expres
7	215	19.3	584	22 AAK34394	Human bone marrow
8	215	19.3	584	22 AAI40117	Probe #8803 used t
9	207.4	18.6	1496	18 AAT69547	Rat pheromone rece

10	205.8	18.5	1053	18 AAT69548	Rat pheromone rece
11	204.4	18.3	1265	18 AAT69550	Rat pheromone rece
12	184.6	16.6	1313	24 AAS98047	Human DNA for pote
13	181.6	16.3	1015	24 AAD24968	Human G-protein co
14	163.2	14.6	628	22 AAH74904	Probe used to Isol
15	158.4	14.2	632	22 AAH74901	Probe used to Isol
16	152.4	13.7	461	22 AAK06912	Human brain expres
17	152.4	13.7	461	22 AAK32640	Human bone marrow
18	152.4	13.7	461	22 AAI38455	Probe #7141 used t
19	151	13.6	632	22 AAH74899	Probe used to Isol
20	149.4	13.4	628	22 AAH74900	Probe used to Isol
21	143.8	12.9	640	22 AAS63212	Human purified sec
22	143.4	12.9	653	22 AAH74903	Probe used to Isol
23	138	12.4	662	22 AAH74902	Probe used to Isol
24	135.4	12.2	636	22 AAH74905	Probe used to Isol
25	134.8	12.1	300	22 ABA72774	Human foetal liver
26	134.8	12.1	300	22 ABA38413	Probe #16879 for g
27	134.8	12.1	300	22 AAK21203	Human brain expres
28	134.8	12.1	300	22 AAK47360	Human bone marrow
29	134.8	12.1	300	22 AAI53196	Probe #21882 used
30	132.2	11.9	574	24 AAS98151	Human DNA for pote
31	128.6	11.5	512	24 ABA05725	Rat pheromone rece
32	127	11.4	512	24 ABA05712	Rat pheromone rece
33	127	11.4	512	24 ABA05731	Rat pheromone rece
34	125.4	11.3	512	24 ABA05704	Rat pheromone rece
35	125.4	11.3	512	24 ABA05705	Rat pheromone rece
36	125.4	11.3	512	24 ABA05710	Rat pheromone rece
37	124.2	11.1	512	24 ABA05721	Rat pheromone rece
38	123.6	11.1	530	18 AAT69551	Human pheromone re
39	120.8	10.8	512	24 ABA05703	Rat pheromone rece
40	120.8	10.8	512	24 ABA05708	Rat pheromone rece
41	119.2	10.7	512	24 ABA05701	Rat pheromone rece
42	119.2	10.7	512	24 ABA05702	Rat pheromone rece
43	119.2	10.7	512	24 ABA05711	Rat pheromone rece
44	119.2	10.7	512	24 ABA05714	Rat pheromone rece
45	119.2	10.7	512	24 ABA05715	Rat pheromone rece

ALIGNMENTS

RESULT	1
AAT69546	
ID	AAT69546 standard; DNA; 1331 BP.
AC	AAT69546;
DF	26-AUG-1997 (first entry)
DE	Rat pheromone receptor VN3 cDNA.
XX	Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW	maternal behaviour; reproductive behaviour; fertility;
KW	hormone secretion; ss.
OS	Rattus sp.
XX	
PH	Key
CDS	Location/Qualifiers
FT	207..1142
FT	/*tag= a
FT	1038..1040
FT	/*tag= b
FT	/transl_except= pos:1038..1040:_aa:Gly
FT	/note= "GTA codes for valine"
FT	1086..1088
FT	/*tag= c
FT	/transl_except= pos:1086..1088:_aa:Glu
FT	/note= "GGA codes for glycine"
XX	
WO9714790-A1.	
PD	24-APR-1997.
XX	

Db 424 gctaagtatcagtaataactgatcacattttctccataagctacattatcagagacctgtc 483
Qy 665 ttccacactaatgactttggaggatgcacacttttataggggttcattggtctctcctcaagagg 724
Db 484 ttcatgtgcactgcttacgaatgtctctttgttagcaatcatcagccctcaagtgc 543
Qy 725 ctacatggtgattttttatcacagaca 751
Db 544 atacatggtaatctctgtttcaggca 570

RESULT 5
ABA28536
ID ABA28536 standard; DNA; 584 BP.
XX
AC ABA28536;
DT 23-JAN-2002 (first entry)
DE Probe #7002 for gene expression analysis in human heart cell sample.
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
XX WO200157274-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1; SEQ ID No 7002; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.38; Score 215; DB 22; Length 584;
Best Local Similarity 64.74; Pred. No. 3e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;
Qy 190 agcaaacaccatctctctttttccacatcttcacctttgttttcagtcacaggtctaa 249

Db 4 agccaaacaacttctctctattcttccaaaattcttctcactcttcaggatcacaggcctaa 63
Qy 250 gtccattgacatgataaattagtcacctgtctctccacacatactgtctctctcaactca 309
Db 64 gccacatgaccccatcaccctgtcaccctggcccttgctccacttaggagtgctccctattgt 123
Qy 310 ggaataattggtgctccttagactctcttgggtccagaaatactcagatgatcttaggta 369
Db 124 ggtcttctggcatctccagacctgtttgagtcactgtattttcagaatgacttcaagtg 183
Qy 370 taaggtcattgtctttttaaacaaaggtgatgagggtcctctccatctgcacccctgcct 429
Db 184 taaggcattctctgcatgacaggggtgatgagggtcctctccatctgcacccctgcct 243
Qy 430 cctgagtgtgtccaggc---catcatcagcccccagcatcttctccttggcacaagctcaa 486
Db 244 cctgagcatgctccaggtgtcgcacatcagccctggcacctcctggtctgcaagattaa 303
Qy 487 acatcctctgcaagtccacatctcttaggattctctcttctctcattggtctcacaactgt- 545
Db 304 acagaaattcaaaaggttacatcttctcactctctcttctctctatgggtctcagttgtc 363
Qy 546 tcattggttaattctctgtgtacactgctgacctacccccagtgaaaacggggccagctct 605
Db 364 tctcagtagtaacctgtctctcactgtggtctcttcttaattgggacaaagacggtgtg 423
Qy 606 -ctgtttgtcatcacagcactgttctctttttgtcccatgagctacacccacagagactgtt 664
Db 424 gctaaagtatcagtaataactgatcactcttctctccataagctacattatcaggagcctgtc 483
Qy 665 ttccacactaatgactttggaggatgtcactcttataccctttataggggtctcctcacaagagg 724
Db 484 ttctatgctgccactgtctacgaatgtctcttctttagcaatcatcagcctcagctcaggtgc 543
Qy 725 ctacatggtgattttttatcacagaca 751
Db 544 atacatggtaatctctctgtttcaggca 570

RESULT 6
AAK08511
ID AAK08511 standard; DNA; 584 BP.
XX
AC AAK08511;
XX
DT 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe SEQ ID NO: 8502.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX

Qy	606	-ctgtttgtcatacagcaactggttccctttttggccatagctacaccacagaggactgtt	664		
Db	424	gctaagtatcagtaaaatactgatcatcttcttctccataagaactaatcatcaggagcctg	483		
Qy	665	tttcacacaataagacttttgagggaatgcaccttttatagggttcattgctctccaagag	724		
Db	484	tttcattgctgcgccactgcttacgaatgtctctcttcttgagcaatcatcgaccctcaagtgc	543		
Qy	725	ctacatgggtgattattttatatcacagaca	751		
Db	544	atacatgggtaattctctgttcaggca	570		
 RESULT 8 AAI40117 ID AAI40117 standard; DNA; 584 BP. XX AC AAI40117; XX DT 17-OCT-2001 (first entry) XX DE DE Probe #803 used to measure gene expression in human placenta sample. XX KW KW Probe; microarray: human; placenta; antenatal diagnosis; KW genetic disorder; ss. XX OS OS Homo sapiens. XX PN PN WO200157272-A2. XX PD PD 09-AUG-2001. XX PF PF 30-JAN-2001; 2001WO-US00663. XX PR PR 04-FEB-2000; 2000US-0180312. PR PR 26-MAY-2000; 2000US-0207456. PR PR 30-JUN-2000; 2000US-0608408. PR PR 03-AUG-2000; 2000US-0632366. PR PR 21-SEP-2000; 2000US-0234687. PR PR 27-SEP-2000; 2000US-0236359. PR PR 04-OCT-2000; 2000GB-0024263. XX (MOLE-) MOLECULAR DYNAMICS INC. XX PI PI Penn SG, Hanzel DK, Chen W, Rank DR; XX WI WI WPI; 2001-48897/53. XX DR DR Human genome-derived single exon nucleic acid probes useful for PT analyzing gene expression in human placenta - XX PS PS Claim 25; SEQ ID No 8803; 654pp; English. XX CC CC The present invention relates to single exon nucleic acid probes (SEN CC The present sequence is one such probe. The probes are useful for CC producing a microarray for predicting, measuring and displaying gene CC expression in samples derived from human placenta. The probes are use CC for antenatal diagnosis of human genetic disorders.					
XX	SQ	Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;			

Query Match	19.3%	Score 215;	DB 22;	Length 584;
Best Local Similarity	64.7%	Pred. No. 3e-50;		
Matches 367;	Conservative	0;	Mismatches 195;	Indels 5;
Gaps 3;				

[illegible]

Qy	310	ggcaataattggtgctcttagactctcttgggttcacagaatactcaggaatgatctttaggta	369
Db	124	ggtctcttggcattctccagacctgtttgagtcactgtatatttcagaatgacttcaagt	183
Qy	370	taaggtcatigtctttttaacaaggtgataggggcctctccatctgcacccctgcct	429
Db	184	taaggcatctcttgcatgcacagggggaagaggagcctctccatctgcacacctgcct	243
Qy	430	cctgagtgctccagc---catcatcagcccccagcatcttctcttggcaagctcaa	486
Db	244	cctgagcatgctccagctgtgccatcagccctggcaccctctgctctgcaaggattaa	303
Qy	487	acatcctcttgcaagtcaacatttaggattcttctcttctcatgggtctccaacatgt-	545
Db	304	acagaaattcaaaaggttacatttctcaactctctctctctcatggttggtctcagttgtc	363
Qy	546	tcattggtgaattctctgtctacactcgtggtacccccagtgaaacggggccagctctt	605
Db	364	tctcagtagtaacctgctctctccactgtggtcttcttaattgggaccaaagacggtgt	423
Qy	606	-ctgtttgttcatacagacactgttctcttttggccatgagctacacccacagagactgtt	664
Db	424	gctaagtatcagtaaatactgatcaactttcttcataagctacattatcaggagcctgc	483
Qy	665	tttcacactaatgactttgaggatgtcacctttatagggttcattggtctctcgaagg	724
Db	484	ttctatgctgccactgcttcacgaatgtctctctttagcaatcatgcagccctcaagtgc	543
Qy	725	ctacatggtgattattttatacagaca	751
Db	544	atacatggttaattctctgttcaggca	570
RESULT 9			
ID	AAT69547	AAT69547 standard; DNA; 1496 BP.	
XX	AC	AAT69547;	
XX	DT	26-AUG-1997 (first entry)	
XX	DE	Rat pheromone receptor VN4 cDNA.	
XX	KW	Pheromone receptor; vomeronasal sensory neuron; social behaviour;	
XX	KW	maternal behaviour; reproductive behaviour; fertility;	
XX	KW	hormone secretion; ss.	
XX	OS	Rattus sp.	
XX	XX	Key	
XX	XX	Location/Qualifiers	
XX	XX	351..1283	
XX	XX	/*tag= a	
XX	PN	W09714790-A1.	
XX	PD	24-APR-1997.	
XX	PF	18-OCT-1996; 96WO-US16637.	
XX	PR	19-OCT-1995; 95US-0005698.	
XX	PA	(UYCO) UNIV COLUMBIA NEW YORK.	
XX	PI	Axel R, Dulac C;	
XX	XX	WPI; 1997-245107/22.	
XX	DR	P-PSDB; AAW19105.	
XX	XX	Nucleic acid molecule encoding vertebrate pheromone receptor .	
XX	PT	useful to identify modulators for control of reproductive and social	
XX	PT	behaviour, fertility and hormone secretion	

PS Claim 1; Fig 10; 123pp; English.

CC cDNA clones (AAT69546-50) respectively code for rat pheromone
CC receptors VN3, VN4, VN5, VN6 and VN7 (AAW19104-08), members of a
CC novel family of presumed 7-transmembrane domain receptors that are
CC evolutionary independent of the odorant receptors of the main
CC olfactory epithelium. These clones, and a clone for VN2 (sequence
CC not provided), were isolated from rat vomeronasal organ cDNA
CC libraries by PCR and hybridisation. A differential cloning method
CC was used to isolate VN7 cDNA (AAT69545). VN1-VN7 (not VN2) clones
CC have been deposited in pBluescript as ATCC 97294-97299. A human
CC homologue, HG25 (AAT69551), has also been isolated. VN polypeptides
CC can be expressed in host-vector systems for use in identifying
CC modulators for control of maternal, reproductive and social
CC behaviour, to increase fertility, control hormone secretion and to
CC regulate food uptake in humans and animals.

XX Sequence 1496 BP; 400 A; 340 C; 282 G; 474 T; 0 other;

Query Match 18.6%; Score 207.4; DB 18; Length 1496;
Best Local Similarity 56.4%; Pred. No. 6.3e-48;
Matches 507; Conservative 0; Mismatches 381; Indels 11; Gaps 6;

QY 118 taagatgatcaaaactctctgtttttattaccatacaaaatcttctttccacatactttttttcag 177
DB 356 taaagacacacactccatgttgacacacatcatgaaatcaactatgtctctgaagtga 415
QY 178 ctttggaatttcagaaacacactctctcttctttccacatactttttttcag 237
DB 416 tgttgcatcttgatcaacagatgattgtttttgttgcactgtgcatgtctcttgaga 475
QY 238 tcacaggtctaaagtcattgacatgataatgatacctgtctcttcacatactgtct 297
DB 476 gaacaagcctaagcccatctcatctctacattgcatctctgtctcttcacacatactgt 535
QY 298 gcttcttaactcaggaacataatgtgtctcttagactcttctgttccacagataactcagga 357
DB 536 gcttataactatggactcatag---ctgctgacatgtttattcttcagggtataggga 592
QY 358 tgatcttaggtataaagtcattctcttttttaacaagagtgatgagggcctctccatctg 417
DB 593 ttctactcatgcagtcctctctatctatcttgacaggtcttcaggggtttacccttag 652
QY 418 caccctcctgctctgagtggtctccaggccatca---tcagccccagcatctctctct 474
DB 653 tgcctgctgtcgtgaagtcttcttgatgacatctcagttcttaaaaaatcctgttt 712
QY 475 gcaaaagctcaaacatctctctgcaagtcacatcttaggattctcttctctcatgggt 534
DB 713 acaaaagtttaaacataaactctcccatcacatctcaggtgcgcttctctctctgtgt 772
QY 535 cctcaacatgt-tcattggtgaattctctgtgtacactgaggtcaccct-cagtgaaa 592
DB 773 tctctacatgtgtttcagcagtcacactattttatctgattgtaccccttaactgac 832
QY 593 cggggccagctctctgtttgtatcacagacactgttctcttttgcccatgagctacacca 652
DB 833 ctcaagataatttatgtattactaagctcgttcttattctaccatgtgttactccag 892
QY 653 caggagactgtttttcacactaactgacttggaggatgtcactttatagggttactggt 712
DB 893 acaaaagctgttttcacacaaactgtgtcagggaagcgttttttatcgtctcatggc 952
QY 713 cctctcaagagctcatggtgattattttatcacagacataaagaggtatctcagtgcc 772
DB 953 cctgtccagtggtgactgtgtgtctctctctgagac-acagggaagcaggcccatc 1011
QY 773 ttcaagcagccagcgtgtcccca---gtctccacagtgaaaagagcctccaggtatct 830
DB 1012 ttcaagcagccagcgtgtgttttttcaagatcatctccagacaaaggccagacattcc 1071
QY 831 tactgctggtgagttttgtcttccacatactgggtggactttacgtttctctcagga 890

DB 1072 tgcgtcttagtgagttcttcttctgtgtctctacatttggaaaatgtgtctctactcaa 1131
QY 891 gtgtgacatgataaaatgattctctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 950
DB 1132 ggaagagttcaagatgagtggtgcaacattctctactgtgtcccaattattgtgtgtgtgtgt 1191
QY 951 atgcccgaattagctctcttctgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 1009
DB 1192 atgcactgtcagctcttctgt 1250

RESULT 10

AAT69548
ID AAT69548 standard; DNA; 1053 BP.

XX AAT69548;

DT 26-AUG-1997 (first entry)

XX Rat pheromone receptor VN5 cDNA.

DE Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW maternal behaviour; reproductive behaviour; fertility;
KW hormone secretion; ss.

OS Rattus sp.

XX Key Location/Qualifiers

FT CDS 3..838

FT /*tag= a

XX W09714790-A1.

XX 24-APR-1997.

XX 18-OCT-1996; 96WO-US16637.

XX 19-OCT-1995; 95US-0005698.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Axel R, Dulac C;

XX WPI; 1997-245107/22.

XX P-PSDB; AAW19106.

PT Nucleic acid molecule encoding vertebrate pheromone receptor -
PT useful to identify modulators for control of reproductive and social
PT behaviour, fertility and hormone secretion

XX Claim 1; Fig 11; 123pp; English.

CC cDNA clones (AAT69546-50) respectively code for rat pheromone
CC receptors VN3, VN4, VN5, VN6 and VN7 (AAW19104-08), members of a
CC novel family of presumed 7-transmembrane domain receptors that are
CC evolutionary independent of the odorant receptors of the main
CC olfactory epithelium. These clones, and a clone for VN2 (sequence
CC not provided), were isolated from rat vomeronasal organ cDNA
CC libraries by PCR and hybridisation. A differential cloning method
CC was used to isolate VN7 cDNA (AAT69545). VN1-VN7 (not VN2) clones
CC have been deposited in pBluescript as ATCC 97294-97299. A human
CC homologue, HG25 (AAT69551), has also been isolated. VN polypeptides
CC can be expressed in host-vector systems for use in identifying
CC modulators for control of maternal, reproductive and social
CC behaviour, to increase fertility, control hormone secretion and to
CC regulate food uptake in humans and animals.

XX Sequence 1053 BP; 281 A; 231 C; 181 G; 360 T; 0 other;

Query Match 18.5%; Score 205.8; DB 18; Length 1053;
Best Local Similarity 58.5%; Pred. No. 1.5e-47;

Matches 470; Conservative 0; Mismatches 322; Indels 11; Gaps 6;		Rattus sp.	
QY 207	ttttttccacatttccacatttgggtttttccaggtcaccaggtcattgacatgataa 266	XX Key	Location/Qualifiers
Db 1	ttttttccacatttccacatttgggtttttccaggtcaccaggtcattgacatgataa 60	FT CDS	138..1061
QY 267	ttagtcacatttccacatttgggtttttccaggtcaccaggtcattgacatgataa 326	FT	/*tag- a
Db 61	ttgcttttccacatttccacatttgggtttttccaggtcaccaggtcattgacatgataa 118	XX	W09714790-AL.
QY 327	tagacttttgggtttccacatttgggtttttccaggtcaccaggtcattgacatgataa 386	PN	24-APR-1997.
Db 119	-agacattgtttatgtctcggggagatggattctaccacatgccagtccttatctatt 177	PD	18-OCT-1996; 96WO-US16637.
QY 387	taaacagggtgatggggcctctccatctgcacccctcctcctcctcctcctcctcag 446	PF	19-OCT-1995; 95US-0005698.
Db 178	tggacaggcttttgggggttttaccctttgtctacctgtctgtaagtcttttggga 237	PR	(UYCO) UNIV COLUMBIA NEW YORK.
QY 447	ccatca---tcagccccagcatttctcctctcctcctcctcctcctcctcctcctc 503	PA	Axel R, Dulac C;
Db 238	ccatcactcactcactcactcactcactcactcactcactcactcactcactcactc 297	PI	WPI; 1997-245107/22.
QY 504	acattctaggattctcttttctcattgggtcctcactcactcactcactcactcact 562	DR	P-PSDB; AAM19108.
Db 298	acattctaggattctcttttctcattgggtcctcactcactcactcactcactcact 357	XX	Nucleic acid molecule encoding vertebrate pheromone receptor -
QY 563	tgtgtgtacactcgggtaccccccagtgaaac-gggggcagctctctgttgttcatacag 621	PT	behaviour, fertility and hormone secretion
Db 358	tttttcaacaattgtaccctcctcctcctcctcctcctcctcctcctcctcctcct 417	XX	Claim 1; Fig 13; 123pp; English.
QY 622	actgtctcttttggccatgagctacacccacacagagagactgtttttcacactaactt 681	CC	cDNA clones (AAT69546-50) respectively code for rat pheromone
Db 418	cctgttatttccacattgatttactcctcagaaacagcatgttttccacacccatggcca 477	CC	receptors VN3, VN4, VN5, VN6 and VN7 (AAM19104-08), members of a
QY 682	tggaggtatcacttctatagggttctcctcctcctcctcctcctcctcctcctcct 741	CC	novel family of presumed 7-transmembrane domain receptors that are
Db 478	tcagggaagcccttcttattggtctcattggtcctgctcagtggtggtacatggttgccttc 537	CC	evolutionary independent of the odorant receptors of the main
QY 742	tatacagacaataagggtctatctcagtccttcacagcagcagcagcagcagcagcag 799	CC	olfactory epithelium. These clones, and a clone for VN2 (sequence
Db 538	tatggagac-acaaagatcaggccggccgctcttcacagcagcagcagcagcagcag 596	CC	not provided), were isolated from rat vomeronasal organ cDNA
QY 800	tcaccagtgaagagcctccaggtctatcttactgtgtggtgtgtgtgtgtgtgtgtgt 859	CC	libraries by PCR and hybridisation. A differential cloning method
Db 597	tcaccagagcaaaaggccacagaccatcatgattctcaggtctgtgtgtgtgtgtgtgt 656	CC	has been used to isolate VN7 cDNA (AAT69545). VN1-VN7 (not VN2) clones
QY 860	tgggtgactttacgttctcctcctcctcctcctcctcctcctcctcctcctcctcct 919	CC	have been deposited in pBluescript as ATCC 97294-97299. A human
Db 657	tacattttggaaaaatgtgtcttctactctcctcctcctcctcctcctcctcctcctc 716	CC	homologue, HG25 (AAT69551), has also been isolated. VN polypeptides
QY 920	gtgtggtcctcaggttatttggcccaatagctatgcgcgaattagctccttgcataatt 979	CC	can be expressed in host-vector systems for use in identifying
Db 717	tactgtgtccaaattattgtgtcccatagctatgcccacatcagcccttttgtgtttatt 776	CC	modulators for control of maternal, reproductive and social
QY 980	tatgtctgataaccaaattatcaa 1002	CC	behaviour, to increase fertility, control hormone secretion and to
Db 777	tgcacagaaagcgtataataa 799	CC	regulate food uptake in humans and animals.
XX		XX	Sequence 1265 BP; 325 A; 287 C; 226 G; 426 T; 1 other;
RESULT 11		XX	
AAAT69550		XX	
ID AAAT69550 standard; DNA; 1265 BP.		XX	
XX		XX	
AC AAAT69550;		XX	
XX		XX	
DT 26-AUG-1997 (first entry)		XX	
TX		XX	
DE Rat pheromone receptor VN7 cDNA.		XX	
XX		XX	
KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;		XX	
KW maternal behaviour; reproductive behaviour; fertility;		XX	
KW hormone secretion; ss.		XX	

Query Match 18.3%; Score 204.4; DB 18; Length 1265;
 Best Local Similarity 55.6%; Pred. NO. 4.1e-47;
 Matches 497; Conservative 0; Mismatches 386; Indels 11; Gaps 5;

QY 479 aagctcaaacatcctctgtgaagtcacatcttaggattctctcttctctatgggtctctc 538
DB 495 aagtttaaacataatccccatcacatctcaggtgcctctctctctctctctctc 554
QY 539 aacatgttcattggtgaattcttc-tgctgacacgtcggtacccccagtgaaacggg 596
DB 555 tatatgctcttagcagtcacctcttggtattggtcattgtctaccccaatttaacctca 614
QY 597 gccagttcttctgttcatacagcactgtctctcttggccatgagctacacccacagg 656
DB 615 gatcattttatgtattgtactcagtcctcactctcaccatgattactcagagaca 674
QY 657 agactgtttttcacacataagctttgaggatgagctcctttataggtttcagtgctc 716
DB 675 agacgttttcttactgattggttcacacagggagagctcttctctatcagtcctcagccctg 734
QY 717 tcaagaggtcacatgggtgattattttatcacagacaataagagactatctcagtgcttca 776
DB 735 tccagtggtgtacatgggtgactctctctctagggc-acaagaagcaggccccagcactctca 793
QY 777 ccagcagcagcctgtc--cccgagctctcaccagtgaaagagcctccaggtctacttact 834
DB 794 cagcaccagactttcttcaaaagcattccccacagcaaaaggccccagaccactctgct 853
QY 835 gctgggtgagtttctctcacactactggtggactttacgttctcatttccagaggtgt 894
DB 854 gcttaggacctcttggtgtttctacatttttaggcactgttatctccactcaaggac 913
QY 895 gacatgataaatgattctctgtctgtagtggctccaggttatgtggccaatagctatgc 954
DB 914 taagtccaagagtggttcaattcttactgtgtcccaattattgttcccatagctatgc 973
QY 955 cgcaattagctctttagctaatattatgctgtgatacccaaatattcaagactct 1008
DB 974 cactatcagccattgtgtgtttgtttttctgaaagcgcataatacaagttttt 1027

RESULT 12

AAS98047
ID AAS98047 standard; DNA; 1313 BP.

XX AAS98047;

AC 12-MAR-2002 (first entry)

XX Human DNA for potential G protein-coupled receptor #5.

KW Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW tuberculosis; cognition disorder; memory disorder; anorexia;
KW hormonal release disorder; cardiovascular activity disorder;
KW pain perception disorder; obesity; diabetes; obesity;
KW diabetes; hyperlipidaemia; stroke; gene therapy.

XX Homo sapiens.

OS

XX WO200185791-A1.

XX 15-NOV-2001.

XX 11-MAY-2001; 2001WO-US15332.

XX 11-MAY-2000; 2000US-203217P.

XX 18-MAY-2000; 2000US-205945P.

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

XX Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

XX

DR WPI; 2002-066595/09.

XX Novel G protein-coupled receptor polypeptides including galanin

PT receptor polypeptides useful for identifying modulators that are useful

PT for treating Alzheimer's disease, psoriasis, melanoma, multiple

XX sclerosis, stroke

XX Disclosure; Page 115; 144pp; English.

XX The invention relates to an isolated polypeptide encoded by a

XX nucleic acid molecule that is at least 80% identical to the G

XX protein-coupled (GPCR) polynucleotides included in the specification.

XX Also included are probes based on the GPCR sequences (including

XX antisense probes), a host cell comprising an expression vector comprising

XX the GPCR sequence, antibodies raised against the polypeptides,

XX and methods of identifying modulators of the polypeptides. The

XX polypeptides are useful for identifying modulator compounds which

XX function as modulators, activators, repressors, agonists or antagonists

XX of the novel GPCR polypeptides including the GAL4 polypeptide. The

XX antibodies and nucleic acid probes as described above can be used to

XX detect the presence of the polypeptides and nucleic acids and are used to

XX diagnose a variety of diseases or disorders in which GPCRs are involved

XX e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma,

XX atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,

XX chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease,

XX depression, epilepsy, macular degeneration, lymphoma, melanoma,

XX multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease,

XX psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis,

XX tuberculosis and many other diseases listed in the specification. The

XX probes and antibodies are also useful for diagnosing cognition and memory

XX disorders, anorexia, hormonal release disorders, cardiovascular activity

XX disorders, pain perception disorders, obesity, diabetes, Alzheimer's

XX disease. Preferably, compounds that decrease or increase

XX the expression of galanin receptor (GAL4) can be used to treat obesity,

XX diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is

XX useful for treating the above mentioned disorders by gene therapy

XX techniques. The present sequence is a novel GPCR polynucleotide of the

XX invention.

XX

XX Sequence 1313 BP; 359 A; 277 C; 206 G; 471 T; 0 other;

Query Match

Best Local Similarity 16.6%; Score 184.6; DB 24; Length 1313;

Matches 496; Conservative 0; Mismatches 314; Indels 43; Gaps 8;

QY 189 cagcaaacaccatcctctctttccacatcttcacattcttcacgttcacaggtcta 248
DB 6 cagcaaacaccatcttctgtttcttcacatcttcacattcttcacgttcacaggtcta 65

QY 249 agtccattgacatgataattagtcacactgtctctccacatactgctcttcactc 308
DB 66 agccacatgactgagtgagcgctacactgtgcttcactgtgagtgagtgctcactg 125

QY 309 aggcaatatgtgctccttagactcttttgggttcacagatactcaggtatgacttaggt 368
DB 126 tagtgattttgtgctccagatattggaatcatttgggaataacttcaaat 185

QY 369 ataaggtcattgtcttttaaacaaagggtgatgagggcctctccatctgacccctgcc 428
DB 186 gcaagtccttgatctataaacagaaatgacgaagggtctatgtttctataccactgtc 245

QY 429 tctgagtggtctccaggc---catcatcagccccagcatcttctctcttggaagctca 485
DB 246 tctgataatacaccagccagcataatacagcctcagcaactctgttggaagcttta 305

QY 486 aacatctcttgcagagtcacattttaggattcttctctt---tctcatgggtcctcaaca 542
DB 306 aacataaatttcaaaaataacattgtcagtgctctcttcttcttcttcttctcctcaatt 365

QY 543 tgttcattggt-gtaattctctgctgtacactgcggtacccccagtg-aaacggggcca 600
DB 543 tgttcattggt-gtaattctctgctgtacactgcggtacccccagtg-aaacggggcca 600

Db	692	atcaatttgcattgctcagtaaaatctgtctcatttttccccataaataaccataatcagaggga	751
Qy	660	ctgtttttccactaatgactttgaggagatgtcaaccttttatagggttcatggtctctctca	719
Db	752	ctgtttttcactctgtcattattcagagatgtttttcttaaacaaaataatgctgtctca	811
Qy	720	agaggctcatggtgattattttatcacagacaataaagagcttatctcagtgctctcaagc	779
Db	812	agtgctccatgatgactctcattcaggaaactacaggag--atcctgggtacccttcacag	868
Qy	780	agccagcctgtcccgcagctccaccagtgaaagag-ctctcccaggctattcttactgctg	838
Db	869	ccccagcctctacctaaaggatctttgcagaggcagaccatcagcacatcctgctgccg	928
Qy	839	gtgagttt---tgtcttcacactacactgggtggaactttacgtttctctcattttcaggagggtgtg	895
Db	929	gtgaggttctcgtgtggcgtgtacaagatggacttcattcattcaacctctcaacgttg	988
Qy	896	acatgcatataatga	909
Db	989	ccatgggcata	1002
RESULT 14			
AAH74904			
ID	AAH74904 standard; DNA; 628 BP.		
XX			
AC	AAH74904;		
DT	29-OCT-2001 (first entry)		
XX			
DE	Probe used to isolate human VNO cDNA sequences. .		
XX			
KW	Human; vomeronasal organ; VNO; pheromone receptor; ion channel;		
KW	pheromone; vomeropherin; probe; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200161046-A2.		
XX			
PD	23-AUG-2001.		
XX			
PF	15-FEB-2001; 2001WO-US05178.		
XX			
PR	17-FEB-2000; 2000US-0183128.		
XX			
PA	(PHER-) PHERIN PHARM INC.		
XX			
PI	Herman RC, Berliner D;		
XX			
DR	WPI; 2001-522609/57.		
XX			
PT	Novel vomeronasal organ cDNA library useful for identification and		
PT	isolation of genes expressed in vomeronasal organs, specifically genes		
XX	for pheromone receptors and ion channels -		
PS	Example 9; Page 54-55; 59pp; English.		
XX			
CC	The present sequence represents a probe, used to isolate human		
CC	vomeronasal organ (VNO) cDNA sequences. The cDNA library was		
CC	constructed from female tissue. Sequences isolated from the VNO cDNA		
CC	library are useful as a unique resource for the identification and		
CC	isolation of genes expressed in VNO specifically genes for pheromone		
CC	receptors, ion channels and prospective reagents for high throughput		
CC	assays. They are useful as an excellent source to search for novel		
CC	genes, gene fragments or other nucleotide sequences encoding proteins		
CC	that are implicated in detection of pheromones or other vomeropherins		
CC	in the human VNO. The cDNAs are useful as hybridization probes for		
CC	determining the presence or concentration of an oligo- or polynucleotide,		
CC	e.g. DNA, of interest.		
XX			
XX	Sequence 628 BP; 138 A; 178 C; 123 G; 189 T; 0 other;		

PA (PHER-) PHERIN PHARM INC.

PI Herman RC, Berliner D;

XX DR WPI; 2001-522609/57.

XX Novel vomeronasal organ cDNA library useful for identification and
PT isolation of genes expressed in vomeronasal organs, specifically genes
PT for pheromone receptors and ion channels -

XX Example 9; Page 52-53; 59pp; English.

XX The present sequence represents a probe, used to isolate human
CC vomeronasal organ (VNO) cDNA sequences. The cDNA library was
CC constructed from female tissue. Sequences isolated from the VNO cDNA
CC library are useful as a unique resource for the identification and
CC isolation of genes expressed in VNO, specifically genes for pheromone
CC receptors, ion channels and prospective reagents for high throughput
CC assays. They are useful as an excellent source to search for novel
CC genes, gene fragments or other nucleotide sequences encoding proteins
CC that are implicated in detection of pheromones or other vomeropherins
CC in the human VNO. The cDNAs are useful as hybridization probes for
CC determining the presence or concentration of an oligo- or polynucleotide,
CC e.g. DNA, of interest.

XX Sequence 632 BP; 150 A; 177 C; 109 G; 196 T; 0 other;

Query Match 14.2%; Score 158.4; DB 22; Length 632;
Best Local Similarity 59.7%; Pred. NO. 2.4e-34;
Matches 358; Conservative 0; Mismatches 231; Indels 11; Gaps 5;

Qy	261	tgataaataacacgtctctcaccacatactcgtctcctcaccacaggcaaatgg	320
Db	1	tgccattggtctctctcctcctcaccacatactcgtctcctcaccacaggcaaatgg	57
Qy	321	tgtcttagactctcttggttcacagatactcagagatgattcttaggtataaaggtcattg	380
Db	58	tagccatagacattttattcttcttgaggaggatggatgacatcatatgtaaatctcttg	117
Qy	381	tccttttaaacagggtgatgaggggcctccatctcgtcacccctcctcctcgtgagtgtc	440
Db	118	tctacttgtagaagtttttagaggtctctctctctctctctctctctcaccctcgtgagtgctc	177
Qy	441	tccaggccatca---tcagcccaagatctctctccttggtggaaagctcaaacatctctctg	497
Db	178	tgcaggccatcacctcagccccaagagctcctgttagcaaaagttcaaacataagctc	237
Qy	498	caagtcacattcttaggattctctctctcctcctcctcctcctcctcctcctcctcctcctc	557
Db	238	cccatcagctctcctctgcccattattctcgtgagcaccctcctcctcctcctcctcctcctc	297
Qy	558	tcctctgctgtacact--gcggctaccccaagtgaaacggggccagctctctctgtttgtca	615
Db	298	acctcttagtatccatcaatgccaccaccaatttgaccacgaacacactttatgcaagta	357
Qy	616	tacagacactgtctcttttggccatgagctacacccacagagagactgtttttcacactaa	675
Db	358	ctcagtcctgctacattacaccttgagttacotcctcctcctcctcctcctcctcctcctc	417
Qy	676	tgactttgagggtgtcaccttttatagggttctcctcctcctcctcctcctcctcctcctc	735
Db	418	tggccatcagagatattctcttattagttctcctcctcctcctcctcctcctcctcctc	477
Qy	736	ttattttacagacaataagagggtctcctcctcctcctcctcctcctcctcctcctcctc	794
Db	478	ttctctgtgtaggc-acagaaatcagatccagcagctctcctcctcctcctcctcctcctc	536
Qy	795	-gagctcaccagtgaaagagcctccagcgtctcctcctcctcctcctcctcctcctcctc	853
Db	537	aaagcatctccagacaagaaggccacacagacacacacacacacacacacacacacacac	596

Search completed: July 21, 2002, 03:57:56
Job time: 7331 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 04:00:29 ; Search time 4107.22 Seconds

(without alignments)
5675.898 Million cell updates/sec

Title: US-09-728-309-2

Perfect score: 1114

Sequence: 1 attccagatcatagatgt.....gagaggtaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_rst.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_rst.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----	-----	-----	-----	-----	-----

1	1098	98.6	206832	2	AL390728	AL390728 Homo sapi
2	1094.8	98.3	156798	30	AL359674	AL359674 Human DNA
3	1094.8	98.3	217087	2	AC104335	AC104335 Homo sapi
4	542.6	48.7	206832	2	AL390728	AL390728 Homo sapi
5	375.6	33.7	217087	2	AC104335	AC104335 Homo sapi
6	313.2	28.1	73661	9	AC074050	AC074050 Homo sapi
7	313.2	28.1	152748	9	AC106730	AC106730 Homo sapi
8	304.6	27.3	65128	9	AC073188	AC073188 Homo sapi
9	304.6	27.3	139384	9	AC006455	AC006455 Homo sapi
10	303.6	27.3	989	4	AB064662	AB064662 Capra hlr
11	281.2	25.2	151941	9	AC108131	AC108131 Homo sapi
12	281.2	25.2	161609	2	AC011823	AC011823 Homo sapi
13	280.6	25.2	95097	9	AC022202	AC022202 Homo sapi
14	279.6	25.1	65128	9	AC073188	AC073188 Homo sapi
15	279.6	25.1	149496	30	AC022860	AC022860 Homo sapi
16	275.4	24.7	194142	9	AC092634	AC092634 Homo sapi
17	268.2	24.7	68299	9	AC091685	AC091685 Homo sapi
18	268.2	24.7	152964	2	AC011821	AC011821 Homo sapi
19	267.2	24.0	909	10	AY065469	AY065469 Mus muscu
20	265.6	23.8	909	10	AY065471	AY065471 Mus muscu
21	262	23.5	178105	9	AC073057	AC073057 Homo sapi
22	255.6	22.9	891	10	AY065468	AY065468 Mus muscu
23	251.6	22.6	903	10	AF291501	AF291501 Mus muscu
24	250.8	22.5	909	10	AY065466	AY065466 Mus muscu
25	249.8	22.4	900	10	AF291499	AF291499 Mus muscu
26	248.4	22.3	161571	9	AC069462	AC069462 Homo sapi
27	248.4	22.3	162339	2	AC024151	AC024151 Homo sapi
28	248.4	22.3	166991	9	AC090884	AC090884 Homo sapi
29	247.6	22.2	918	10	AY065467	AY065467 Mus muscu
30	246.2	22.1	1020	10	AY065473	AY065473 Mus muscu
31	244.8	22.0	188389	2	AC024199	AC024199 Homo sapi
32	244.8	22.0	195905	2	AC092447	AC092447 Homo sapi
33	244.4	21.9	894	10	AY065477	AY065477 Mus muscu
34	244.4	21.9	918	10	AY065470	AY065470 Mus muscu
35	242.6	21.8	152964	2	AC011821	AC011821 Homo sapi
36	241.2	21.7	912	10	AY065489	AY065489 Mus muscu
37	240.6	21.6	891	10	AY065474	AY065474 Mus muscu
38	240.4	21.6	202333	2	AC080144	AC080144 Mus muscu
39	239.6	21.5	1738	9	AF253312	AF253312 Homo sapi
40	239.4	21.5	957	10	AB062896	AB062896 Mus muscu
41	238.8	21.4	3403	10	MMPEREC2	Y12734 Mus musculu
42	238	21.4	897	10	AY065487	AY065487 Mus muscu
43	237.6	21.3	912	10	AY065488	AY065488 Mus muscu
44	237.4	21.3	900	10	AY065486	AY065486 Mus muscu
45	236.2	21.2	223275	2	AC098739	AC098739 Mus muscu

ALIGNMENTS

RESULT 1

AL390728

LOCUS

DEFINITION

PROGRESS

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL390728 206832 bp DNA linear HTG 06-FEB-2002
Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
AL390728
AL390728 22 GI:18614040
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Ashwell, R.
Direct Submission
Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 7, 2002 this sequence version replaced gi:18491329.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: bA488LJ8
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Et-amerham; 9% of reads Chemistry:
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.91x in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.
 Location/Qualifiers
 1..206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone_lib="RP11-488L18"
 /clone_lib="RP11-488L18"
 /clone_lib="RP11-488L18"
 /note="assembly_fragment:03301"
 misc_feature
 1..204689
 /note="assembly_fragment:03301"
 misc_feature
 204790..206832
 /note="assembly_fragment:03801"
 BASE COUNT 58135 a 44577 c 44388 g 59631 t 101 others
 ORIGIN
 Query Match 98.6%; Score 1098; DB 2; Length 206832;
 Best Local Similarity 99.7%; Pred. No. 3e-273;
 Matches 1111; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
 QY 1 attccagatcatagagatgttgaaattggtattattgagaaacatgcagaaattatgct 60
 Db 144680 ATTCACATCATAGAGATGTTGAATTTGGTTATTATTGAGAACATGCGACAAATATATGCT 144739
 QY 61 attctattagatctctgtttttctccacagatatcttctgttttaatttctcttaa 120
 Db 144740 ATTCATATTAGATCTCTGCTTTTCTCCACAGATATCTTTGCTTTAATTTTCTCTTAA 144799
 QY 121 gatgatacaacttctctgtttttattacatacaaaattcttcttttatccaaagccagctt 180
 Db 144800 GATGATCAAACTTCTCTGCTTTTATTACCATACAAATCTTCTTTATCCACAGCCAGCTT 144859
 QY 181 tggatttcagcaaacaccatctcttctttccacatcttccacttttttttttttttttttttt 240
 Db 144860 TGGAAATTCAGCAAAACACCATCTCTCTTTCCACATCTTACCTTTGTTTTCAGTCA 144919
 QY 241 caggtctaaagtcattgacatgataattagtcacgtgtctctctatccacatactgtgct 300
 Db 144920 CAGGTCTAAAGTCCATTGACATGATAATTAGTCACCTGTCTCTCATCCACATCTGCTGCT 144979
 QY 301 ctccactcaggcaataattggtgtctcttagacttttttttttttttttttttttttttttttt 360
 Db 144980 CTTCACATCAGGCAATATTTGGTCTCTTTAGACTCTTTTGGTTTCAGAAATACTCAGGATGA 145039
 QY 361 tcttagtataaggtcattgtcttttttaacaaggtgataggggctctccatctgac 420
 Db 145040 TCATTAGGTATAGGTCTATTGCTTTTAAACAAGGTGATAGGGGCTCTCCATCTGCAC 145099

QY 421 cccctgctctcagtgctccaggccatcatcagcccgagcatcttctctctggcaaa 480
 Db 145100 CCCCTGCTCTCTGAGTGTGTCTCCAGGCCATCATCAGCCCGCAGCATCTTCTCTGGGCAAA 145159
 QY 481 gctcaaacatctcttctgcaagtcacatcttaggattctctctctctctctctctctcaaa 540
 Db 145160 GCTCAAAACATCTTCTGCAAGTCACATCTTAGGATTCTTCTCTCTCTCTCTCTCAAA 145219
 QY 541 catgttcatgggttaattctctgctgtacactgctgctacccccccagtgaaaggggcca 600
 Db 145220 CATGTTCAATTTGGTGAATCTTCTGTACACTGCGCTACCCCCAGTGAACGGGCCA 145279
 QY 601 gtctctgttctcatcacagcactgttctcttttggccatgagctacacccacagagac 660
 Db 145280 GTCTCTGTTTGTTCATACAGCAGCTGTTCCCTTTTGGCCATGAGCTACACCCACAGGAGAC 145339
 QY 661 tgttttcaactaactgactttgagggatgctacacctttatagggttcatggtctctcaaa 720
 Db 145340 TGTTTTTCACACTAATGACTTTGAGGGATGTCACCTTTATAGGGTTTCTATGCTCTCTCAAA 145399
 QY 721 gagctacatgggtattattttatatacagacaataaagaggtatctctcagtccttcacaca 780
 Db 145400 GAGCTACATGGGTGATTTATTTATACAGACAATTAAGAGGCTATCTCAGTGCCTTCACACA 145459
 QY 781 gccagcctgtcccgagctctcaccagtgaaagagcctccaggtctatcttactgctggt 840
 Db 145460 GCCAGCCTGTCCCGAGCTCTCACCCAGTGAAGAGAGCCTCCAGGCTATCTTACTGCTGTT 145519
 QY 841 gagtttcttcaactactggttgagcttttactgcttctctctctctctctctctctctcaatg 900
 Db 145520 GAGTTTGTCTTTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 145579
 QY 901 gataaatgattctctgctagtgctgctccaggttattgtggtggtggtggtggtggtggtggt 960
 Db 145580 GATAAATGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 145639
 QY 961 tagtctttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020
 Db 145640 TAGTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 145699
 QY 1021 gtttaaatatttctctctccaaagctcatgttgaaatttaatttaatttaatttaatttaatt 1080
 Db 145700 GTTTAAATATTGCTCTCTCTCCAAAGCTCATGTTGAAATTTAATGCGCAATGTGCGCAGTAC 145759
 QY 1081 taagaagtgatgagaggttaattccattcatg 1114
 Db 145760 TAAGAAG---TGATGAGAGGTTAATCCATTTCATG 145790

RESULT 2

AL359674/C

ID AL359674 standard; DNA; HTG; 166798 BP.

XX AC AL359674;

XX SV AL359674.4

XX DT 20-JUN-2000 (Rel. 64, Created)

XX DT 20-JAN-2001 (Rel. 66, Last updated, Version 5)

XX DE Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-433K2

XX HTG; HTGS_DRAFT; HTGS_PHASE1.

XX OS Homo sapiens (human)

XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

XX XX Eutheria; Primates; Catarrhini; Hominidae; Homo.

XX RN [1]

XX RA Plumb B.;

XX RT ;

XX RL Submitted (20-JAN-2001) to the EMBL/GenBank/DBJ databases.

Db 153354	CCGAGCTGTCCCGAGTCTCACACAGTGAAGAAGAGCCTCCAGAGCTATCTTACTGCTGGT	153295
Qy 841	gagtttgccttcacatactacggtggtgagcttttaacgtttcattcttcattttcaaggagdggtgacatg	900
Db 153294	GAGTTTGTCTCTTCACATACTGGGTGGAGCTTTTACGCTTCTCTCAATTTTTCAGGAGGTGTGACATG	153235
Qy 901	gataaagattctctgctagtgctgctccagggttatgtgccaatagctatgcgcgaat	960
Db 153234	GATAAATGATTCTCTGCTAGTGTGGCTCCAGGTATTGTGGCCAATAGCTATGCGGCAAT	153175
Qy 961	tagtcccttgcgtctaatttatgcgtgataaccacaatttcaagactctgcaaatgtttatg	1020
Db 153174	TAGTCCCTTTGATGCTAATTTATGCTGATACCAAAATATTCAAGACTCTGCAAAATGTTTATG	153115
Qy 1021	gtttaaatattgtctctccaaagctcatgttgaaatttaattgccaatgtggcagtagtac	1080
Db 153114	GTTTAAATATTGTTGCTCTCTCAAAAGCTCATGTTGAAATTTAATGCCAATGTGCAGTAC	153055
Qy 1081	taagaagtgcgtgagtgaggttaattccattcatg	1114
Db 153054	TAGAAG---TGATGAGAGGTTAATCAATTCATG	153024
RESULT 3		
AC104335/c		
LOCUS	AC104335	217087 bp DNA linear HTG 08-DEC-2001
DEFINITION	Homo sapiens chromosome 1 clone RP11-433K2, WORKING DRAFT SEQUENCE,	
	6 unordered pieces.	
ACCESSION	AC104335	AL359674
VERSION	AC104335.1	GI:17426299
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 217087)	
AUTHORS	Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.	
JOURNAL	Direct Submission	
REFERENCE	2 (bases 1 to 217087)	
AUTHORS	Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-DEC-2001) Genome Center, University of Washington,	
	Box 352145, Seattle, WA 98195, USA	
COMMENT	On Dec 8, 2001 this sequence version replaced gi:9801069.	
	----- Genome Center	
	Center: University of Washington Genome Center	
	Center Code: UWGC	
	Web site: http://www.genome.washington.edu	
	Contact: uwgchgs@u.washington.edu	
	Drafting Center: SC	
	----- Project Information	
	Center project name: chr-1	
	Center clone name: RP11-433K2 (sc0158)	
	----- Summary Statistics	
	Sequencing vector: plasmid; 50% of reads	
	Sequencing vector: plasmid; L08752; 50% of reads	
	Chemistry: Dye-terminator ET; 64% of reads	
	Chemistry: Dye-terminator Big Dye; 36% of reads	
	Assembly program: Phrap; version 0.990319	
	Consensus quality: 213380 bases at least Q40	
	Consensus quality: 215084 bases at least Q30	
	Consensus quality: 216045 bases at least Q20	
	Insert size: 256066; 17.2% error; agarose-fp	
	Insert size: 216587; sum-of-contigs	
	Quality coverage: 6.3x in Q20 bases; agarose-fp	
	Quality coverage: 7.5x in Q20 bases; sum-of-contigs	

	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 6 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4074: contig of 4074 bp in length
 * 4075: gap of unknown length
 * 4175: contig of 5534 bp in length
 * 9709: gap of unknown length
 * 9809: contig of 3834 bp in length
 * 13642: gap of unknown length
 * 13643: contig of 39104 bp in length
 * 13743: gap of unknown length
 * 52847: contig of 87484 bp in length
 * 52947: gap of unknown length
 * 140431: contig of 76557 bp in length.
 * 140531: 217087: contig of 76557 bp in length.

FEATURES

source

Location/Qualifiers
 1. .217087

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-433K2"

/clone_lib="RPC1 human BAC library 11"

misc_feature

/note="assembly_name:Contig168"

misc_feature

4175..9708

misc_feature

/note="assembly_name:Contig169"

misc_feature

/note="assembly_name:Contig170"

misc_feature

/note="assembly_name:Contig171"

misc_feature

52947..140430

misc_feature

/note="assembly_name:Contig172"

misc_feature

140531..217087

misc_feature

/note="assembly_name:Contig173"

misc_feature

60435 a 48981 c 48285 g 58724 t 662 others

BASE COUNT

ORIGIN

Query Match 98.38; Score 1094.8; DB 2; Length 217087;
 Best Local Similarity 99.68; Pred. No. 2e-272;
 Matches 1109; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

Qy 1 attccagatcatagagatgttgaaattgggtattattgagaaacatggcagaaattatgct 60
 Db 121154 ATCCAGATCATAGAGATGTTGAAATGGTTATTATTGAGAAACATGCGAGAAATATGCT 121095
 Qy 61 attctcattagatctctgtctttctccacagatatactctgtcttttaattctctctaa 120
 Db 121094 ATTCTCATTTAGATCTCTGCTTTTCTCCACAGATATCCTTTGCTTTAATTTCTCTTAA 121035
 Qy 121 gatgatcaaacctctcgtgtttattaccatacaaatctcttttatccacagcagctt 180
 Db 121034 GATGATCAAACTTCTCGTGTATTATACCATATAATCTCTTTTATCCACAAAGCCAGCTT 120975
 Qy 181 tggaaattcagcaaacacacacctctcttctttccacatcttcacacctgtgttttcagtca 240
 Db 120974 TGGAAATTCACAAACACCATCTCTCTTTTCCACATCTTCACTTTGTTTTCAGTCA 120915
 Qy 241 caggtctaagtccaattgacatgataattagtcacactgtctctcaccacatactgctgct 300
 Db 120914 CAGGTCTAAGTCCATTGACATGATAAATTAGTACGCTGTCTCTCATCCACATACCTGCTGCT 120855
 Qy 301 ctccactcaggaacaattgtgtctcttagactctcttggtttccacagatactcagatga 360
 Db 120854 CTTCCACCAGGAATATGTTGTCCTTAGACTCTTTGGTTTCAGATACTCAGGATGA 120795
 Qy 361 tcttaggtataaggtcattgtcttttttaacaaggtgatgaggggcctctccatctgcac 420
 Db 120794 TCTTAGGTATAAGTTCATTGCTCTTTTAAACAAGTGATGAGGGCCTCTCATCTGCAC 120735
 Qy 421 cccctgctctcgtagtgctccaggccaatcagccccagcatctctctcttggcaaa 480
 Db 120734 CCCCTGCTCTCTGAGTGTCTCCAGGCCATCATCAGCCCCCAGCATCTCTCTCTGGCAAA 120675

Qy 481 gctcaaacatctctctcgaagtccacatctttaggattcttctcttcttctctctctcaaa 540
 Db 120674 GCTCAAAACATCTCTTCTGCAAGTCACATCTTAGGATCTTCTCTTCTCTCATGGTCTCA 120615
 Qy 541 catgttcattggtgtaattctctctgtacactgctgggtacacccagtgaaacggggcca 600
 Db 120614 CATGTTCAATGGTGTAAATCTTCTGTGTACACTGCTGGGTACCCCGAGTGAACGGGGCCA 120555
 Qy 601 gtctctgtttgttcatacagcaactgttctcttcttcttcttcttcttcttcttcttctt 660
 Db 120554 GTCTTCTGTTTGTATACAGCACTGTTCCTTTTGTGCCCATGAGCTACACCCACAGGAGAC 120495
 Qy 661 tgttttcacactaatgacttttgaggatgtcaccttttatagggttcatgtctctctcaaa 720
 Db 120494 TGTTTTTCACACTAATGACTTTTGAGGATGTACACTTTATAGGGTTTCATGGTCTCTCTCA 120435
 Qy 721 gaggtacatggtgattattattttatcacagacaataagaggctatctcagtcgcttcacaca 780
 Db 120434 GAGGCTACATGGTGATTATTTTATACAGACAATAAGAGGCTATCTCAGTGCCTTTCACACA 120375
 Qy 781 gccagcctgtcccagagctctccacagtgaaaagagctcccccagcgtctcttactgctggt 840
 Db 120374 GCCAGCCTGTCCCGAGTCTCACCAGTGAAAAGAGCCCTCCCGAGGCTATCTTACTGCTGT 120315
 Qy 841 gagtttgttttcacatactactggtgagctttacgtttctcatttccagagagtgagacatg 900
 Db 120314 GAGTTTGTCTTCACATACTACTGGGTGGACTTTTACGTTCTTCATTTTCAGAGGTTGTGACATG 120255
 Qy 901 gataaatgattctctgctagtggtgctccaggtattattgtgcccataagctatgcgcgaat 960
 Db 120254 GATAAATGATTCTCTGCTAGTGTGCTGCCAGTTATTGTGCCAATAGCTATGCCGCAAT 120195
 Qy 961 tagtctttgagtgataatttatgtgataccaaataatcagaagctctgcaaatgttatg 1020
 Db 120194 TAGTCCTTTGATGCTAAATTTATGCTGATAACCAATATTCGAAGACTCTGCAAAATGTTATG 120135
 Qy 1021 gtttaaatattgtctctccaaagctcatgttgaatttaattcgcgaatgtgcagttac 1080
 Db 120134 GTTTAAATATTGTTCTCTCTCCAAAGCTCATGTTGANAATTTAATGCCCAATGTGGCAGTAC 120075
 Qy 1081 taagaagtgatgatgagagggttaattccattcatg 1114
 Db 120074 TAAGAAG---TGATGAGAGGTTAATCCATTTCATG 120044

RESULT 4

AL390728/c

LOCUS

DEFINITION

Homo sapiens

AL390728

ACCESSION

AL390728

VERSION

HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (sites)

Direct Submission

Ashwell, R.

TITLE

Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 7, 2002 this sequence version replaced gi:18491329.

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA488L18

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; 50% of reads
 Chemistry: Dye-terminator Er; 64% of reads
 Chemistry: Dye-terminator Big Dye; 36% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 213380 bases at least Q40
 Consensus quality: 215084 bases at least Q30
 Consensus quality: 216045 bases at least Q20
 Insert size: 256066; 17.2% error; agarose-fp
 Insert size: 216587; sum-of-contigs
 Quality coverage: 6.3x in Q20 bases; agarose-fp
 Quality coverage: 7.5x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4074: contig of 4074 bp in length
 * 4075 4174: gap of unknown length
 * 4175 9708: contig of 5534 bp in length
 * 9709 9808: gap of unknown length
 * 9809 13642: contig of 3834 bp in length
 * 13643 13742: gap of unknown length
 * 13743 52846: contig of 39104 bp in length
 * 52847 52946: gap of unknown length
 * 52947 140430: contig of 87484 bp in length
 * 140431 140530: gap of unknown length
 * 140531 217087: contig of 76557 bp in length.

FEATURES

source

1. 217087
 /location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-433k2"

/clone_lib="RPC1 human BAC library 11"
 1. 4074

misc_feature

/note="assembly_name:Contig168"

misc_feature

4175..9708
 /note="assembly_name:Contig169"

misc_feature

9809..13642
 /note="assembly_name:Contig170"

misc_feature

13743..52846
 /note="assembly_name:Contig171"

misc_feature

52947..140430
 /note="assembly_name:Contig172"

misc_feature

140531..217087
 /note="assembly_name:Contig173"

BASE COUNT 60435 a 48981 c 48285 g 58724 t 662 others

ORIGIN

Query Match 33.7% Score 375.6; DB 2; Length 217087;
 Best Local Similarity 77.9% Pred. No. 2.5e-86;
 Matches 547; Conservative 0; Mismatches 134; Indels 21; Gaps 7;
 Qy 1 attccagatcatagaga-----tggtgaaattggtttattgagacatggcagaaa--- 53
 Db 139231 ATCCAGATCAGACAGATTATTAACATAATGATACATAATTAACACATGCGCACAATGT 139290
 Qy 54 ---ttatgctattctattagatctctgtctttctccacagatattcttggcttaatt 110
 Db 139291 TCCTAATGCAATATCATCTTTGATGCTCTATTATTTCTCTACAGATATCCTTTATTGTTT 139350
 Qy 111 ttcctttaag---atatacaaaattcctggtttttattacatacaaaattctttatc 167
 Db 139351 AATTTTCATGGAATGACCAAGCTTTTCAGCTATATGTCATTAATAAATGTTTATTACC 139410
 Qy 168 cacaagccagcttggatttcagacaacaccatctcttctttccacatcttcacct 227

Db 139411 CTCAACTCAGCTTTGGAAATCTCAGCAAAACACCTTCCTCCTCTCTTTTCCACATCTTCACTT 139470
 Qy 228 ttgttttcagtcacaggtcttaagtcctattgacatgataaattagtcacctgtctctcatcc 287
 Db 139471 TTGCTTATACTTCACAGGCTTAAGCCCATTTGACATGACCATAGTACTAGTCTGCCCCCTAATCC 139530
 Qy 288 acatactgctgctcttcactcaggcaatatattggtgctcttagaacttttttggttcaaga 347
 Db 139531 ACATACTGCTACTCTTCACTCAGGCGATACCTGGTCTCTCCGACTTATTGGAATCATGGA 139590
 Qy 348 atactcag---gatgatcttaggtataggtcattgtctcttttaaaagaaggtgatgagg 404
 Db 139591 ATATTGAGAACATGATCTAAAGTGTAAAGATCATCATCATTTTAAACAGGGGTGATGAGG 139650
 Qy 405 gctctccattcgcacccccctgctcctctgagtgtgctccaggccat---catcagcccca 461
 Db 139651 GAGTCTCCATCTGCACCACCTTGCCTCCTTAAGCGTGTCCAGGCCATCACCATCAGCCCCA 139710
 Qy 462 gcattctctcttggaagaagctcaaacatctcttctgaaagtcacatcttagagattcttcc 521
 Db 139711 GCACCTCCTTCTTGGAAAGTTTAAACATATTTCTGCAAAATCACACCTTAGGCTTCATCC 139770
 Qy 522 ttctcatgggtcctcaacatgttcatlgtt-gtaatcttctgctgtacactgcggcta 580
 Db 139771 TCTTCTCATGGGCTCTCAACATGTTCACTAATAACCTTTTGTCTTTTCATTTGTGCTA 139830
 Qy 581 cccccagt-gaaacggggccagctctctgtttgtcacaacagcactgttcttcttttgcca 639
 Db 139831 CCCGGAATAGGATTGGGCCAGTCTTTTGTGTGTCACCTGAGCAGCTGTTATGTTTGGCCA 139890
 Qy 640 tgagctacacacacagagagactgtttttcacacataagactt 681
 Db 139891 TGAGCTACACCACAGGAGCCTGTTTTTCATACATAAAGCCTT 139932

RESULT 6

AC074050/c

LOCUS AC074050 73661 bp DNA linear PRI 04-JAN-2002
 DEFINITION Homo sapiens chromosome 16 clone CTD-2358C21, complete sequence.
 ACCESSION AC074050
 VERSION AC074050.3 GI:18057079

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 73661)

AUTHORS

DOE Joint Genome Institute.

TITLE

Sequencing of Human Chromosome 16

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 73661)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (13-JUL-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 73661)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Jan 4, 2002 this sequence version replaced gi:15963644.

Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality

Oy	784	agctgtcccc--gagtctcacagtgaaaagagcctcccccaggcttatcttactgctggtg	841
D6	72858	CACCTCTCCCCAAGACCCTCCCGAGAAAAAGGGCCACCCAGACCATCCTGCTGCTGTG	72799
Oy	842	agt---tttgtcttcacatactgggtgaccttttacgttctcatcttcaggaggtgttgaca	898

Qy	899	tggaataagattctc---tggctagtggtccagggtattgtggccaatagctatgcc	955
Db	72738	TGGCGGTATGACCCACACGTTGCTCTGAGTGTGCAGAGGCTGTGCTCAATGCCTATGCC	72679
Qy	956	gcaattagctccttgatgctcaattattatctctgataccaacaattattcaagactctgcaaa	1013
Db	72678	ACTGTTAGTCTCTTTGGTCAAAATCAGTTCTTGATAAACAATAATCAGTATTATGCAAA	72621
RESULT	7		
AC106730			
LOCUS	AC106730	152748 bp	DNA linear PRI 03-FEB-2
DEFINITION	Homo sapiens chromosome 16 clone CTD-2014E2, complete sequence.		
ACCESSION	AC106730		
VERSION	AC106730.2	GI:18482002	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 152748)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Sequencing of Human Chromosome 16		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 152748)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, U		
AUTHORS	3 (bases 1 to 152748)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, U		
AUTHORS	On Feb 3' 2002 this sequence version replaced gi:18139280.		
TITLE	Sequence Quality Assessment:		
JOURNAL	This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.		
COMMENT	All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.		
	Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.		

	Sequence Quality Assessment:		
	This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.		
	All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.		
	Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.		

FEATURES	Location/Qualifiers		
source	1. 152748		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/clone="CTD-2014E2"		
BASE COUNT	46721 a 30969 c 30999 g 44059 t		
ORIGIN			

```
Query Match      28.1%; Score 313.2; DB 9; Length 152748;
Best Local Similarity 62.8%; Pred. No. 3.4e-70;
Matches 639; Conservative 0; Mismatches 358; Indels 21; Gaps 9;

Oy 15 agatgttgaaattggttattattagacacgagcagaattattctcttattcattatgc 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132487 AAAAGATGATGATGTAACAATTTTGAACACTGAACACCAACCAATTTGATAACATGGGATC 132546

Oy 75 tcttgctt-----ttccacagatatcccttctttaaatttcttcccttaagtatgaca 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132547 TATTGTTCTTTTATTCAGATATAGAAAGTATGTTCTCTATGCCATTAAGATGAAC 132606

Oy 129 aactcttggtttattaccatacaaatctctttttatcccaagcagctttggaattt 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132607 TATTTTCTACTTATGCTATCCTTAAAAATGCTTTGTTTTCAAAGCTGGCATTTGCTTCT 132666

Oy 189 cagcaaacaccatcctctctcttccacatctccacttctgcttcttcttcttcttcttca 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132667 CAGCCACACCTTTCTCTCTTTCTCCACATCTTTAGCTTCTTCTGAATCGCAGGCTTA 132726

Oy 249 agtcattgacatgataattagtcacactgtctctctccacacatactctgtctcttctcactc 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132727 AACCCCGTGACTTACTACCTGTTCACCTGGCCCTCATTCACATTCAGATTCCTCTCACTG 132786

Oy 309 aggcaatatgtgtctcttagaactcttcttcttcttcttcttcttcttcttcttcttcttctt 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132787 CAGTGGGATTTTTCGCTCTAGACATATTTGAATCACTGCAATTTTGGGAATGACTTTTAAGT 132846

Oy 369 ataaggttcattgtctttttaaacaaggtgtagagggcctctccatctgcagccccctgcgc 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132847 GTAAGGCATTTTATACACAACAGGGCAATGAAGGGCCTCTCCATCTGCACACCTGCC 132906

Oy 429 tcttgagtgctcagggccat----catcagccccagcatcttctctcttcttcttcttcttcttca 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132907 TTTCTGAACATGCTCCAGGCCATTCAGCATTCAGCCCGCAGCACCTCTCTGGTTGGCAAGGTTTA 132966

Oy 486 aacatctctcgaagtcacatcttagattcttctcttcttcttcttcttcttcttcttcttcttca 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132967 AACAATAATCCACAATATACATTTTGCAATGTTTCTTCTTTGGGTGTCCTCAATTTGT 133026

Oy 546 tcattggt-gtaattctctgtctacactcggtcagctacccccagtgaaacgggg--ccagtc 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133027 CTTTCAGTGTAGTTGGCCAATCTTTTCTGTGTGATCTTCTCCACACGACCCATATCAAGC 133086

Oy 604 tctgtttgttcacacagactgttctcttcttcttcttcttcttcttcttcttcttcttcttcttct 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133087 TACAGAATGTCAGTAAATFACTGCCACTTTC--CCCCATGAACCTCCACAAATCAGGGGAGTGT 133145

Oy 664 ttctcaactaatgactttgaggtgtcacctttatagggttcatggtctctctcaagag 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133146 TTTTCACTGTGACATTTATCCAAGTATGTCCTGTGTAGTACTTATGCTGTCTCAAGGG 133205

Oy 724 gctacatggtgattattttatcacagacaataaaggaggtctatctcagtgcttccacacagcc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133206 CATACATGTTGATTTCTCTGTCGCCAGGC--ATTGGAGGAATAATCCAGCACCTCATAGCAC 133264

Oy 784 agcctgtcccc--gagtcctcaacagtgaaagagcctcccaaggtctattctactctggtg 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133265 CACCTCTCCCCAAGACCGCTCCCCAGAGAAAAGGGCCACCCAGACCATCTCTCTCTGCTG 133324

Oy 842 agt---tttgtcttcacatactcgtgtggaactttacgttctctcttcttcttcttcttcttcttca 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133325 AGTTTCTTTGTCGTCAGTACTGGGTGGACATTTATCTTCTCAACCTCTTTAAACCTGCTA 133384

Oy 899 tggataaatgattctc---tgtctagtggtgctccaggttatttgggccaatagctatgcc 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133385 TGGCGGTATGACCCACACAGTGTCTCTGAGTGTGCAGAGAGCTGTGCTCAATGCCTATGCC 133444

Oy 956 gcaattagtccttctgactgaataattatgctgataacacaaatattcaagactctgcaaa 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133445 ACTGTTAGTCTCTTTGGTGAATCAGTCTCTGATATAAAACAATAATCAGTATTATGCAAA 133502
```

RESULT 8
AC073188/c
LOCUS
DEFINITION
AC073188
AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188 65128 bp DNA linear PRI 09-JAN-2002
Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

AC073188.10 GI:15920117
HTG.
human.
SOURCE

donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

```

source
1..65128
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7"
/clone="RP11-73B2"
/clone_lib="RPC1-11"
406..1310
/rpt_family="L1"
1317..1547
/rpt_family="L1"
3257..3634
/rpt_family="L1"
3635..3925
/rpt_family="MaLR"
3930..4120
/rpt_family="MaLR"
4136..4334
/rpt_family="Alu"
4341..4488
/rpt_family="Alu"
4555..4753
/rpt_family="L1"
5597..5693
/rpt_family="MER1_type"
5802..6205
/rpt_family="L1"
9188..9281
/rpt_family="MER53"
10528..10758
/notes="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="(TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/notes="similar to EST AI806860 (NID:g59393426) wf36b07.x1"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="(TTTTTA)n"
14129..14883
/rpt_family="L1"

```

```

repeat_region 15223..15246
/rpt_family="AT_rich"
misc_feature 15290..15571
/notes="similar to EST AW445087 (NID:g6986849)"
misc_feature 15290..15571
/notes="similar to EST AW627961 (NID:g7374751) hi36d12.x1"
repeat_region 16305..16574
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="(CAAAA)n"
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="(A)n"
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="(GAAA)n"
19256..19691
/rpt_family="ERV1"
19748..19790
/rpt_family="(CAAAA)n"
20557..20613
/notes="similar to EST AW445087 (NID:g6986849)"
20557..20666
/notes="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
20594..20828
/notes="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/notes="similar to EST BF899282 (NID:g12290741)"
22640..23050
/notes="similar to EST AI279713 (NID:g3917947) ql53f03.x1"
22855..23086
/notes="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
23402..23451
/rpt_family="T-rich"
23430..23733
/rpt_family="Alu"
23850..24148
/rpt_family="MER2_type"
24566..25150
/rpt_family="L1"
25151..25275

```

Query Match 27.3%; Score 304.6; DB 9; Length 65128;

Best Local Similarity 62.1%; Pred. No. 5.4e-68;

Matches 639; Conservative 0; Mismatches 364; Indels 26; Gaps 9;

QY 11 atagagatgttgaaatgggtattattgagaacatggcagaaattatgctattctcatta 70

Db 37223 ATTAATAATGATAATATGGTAATAATTTTGACCGTGAATAATAATACTACTGATGATAGGA 37164

QY 71 gatctctgtttttctccacagatatcctttgttttaatttttcttcttaagatcaaaa 130

Db 37163 TATCTATTGCTTTTTCAGATATATAAAATATGATCCCTCTTCCACATTAAGACGACCATTA 37104

QY 131 ctctctgtgttttattaccacaaattctttttatccaaagccagcttttgggaattcca 190

Db 37103 G-TCCACTTATGACATATTATAAAATTTGCCCTTTTTCCTCCCAAGCTGGTATTGGACTCACA 37045


```
QY 714 ctctcaagaggtactagtggtgatttattttatcacagacaataaaggaggtatctcagtgctt 773
Db 601 CTCCTCAAGTGGATACATGGTGACTCTCTGTACAGGC-ATAAGAGGCAGACCCAGCATCT 659
QY 774 tcacacagcagcctgtccccc--gagtcctcacagtgaaagagcctccacaggctatctt 831
Db 660 TCAGAGCACCAGCGCTTCTCCAAAAGCATCCCAAGAAAGAGGGCCACCCAGAGCATCTCT 719
QY 832 actctggtgagt--ttgtcttcacatactgggtgagctttacagtttctcattttcagg 888
Db 720 ATTCTCTCATGAGTCTCTTTGGTGCATGTACTTTTGGACATAACTGTCTCTTGGTCTC 779
QY 889 aggtgtgacatgataaattgattctctgctagtgctgctccaggttattgtggccaatag 948
Db 780 AAGATGTTGGGATATCGACTCTGTTCGTATGTGTGTCAGATGCTTGTGGGAATGG 839
QY 949 ctatgcgcgaatagtcctttgtatgactaatttatgtctgataacccaataattcaagactct 1008
Db 840 CTATGCCAGCATCAGCCCTTTGGTGCATCTGAATCAGTACTGAAAAGAGAATAATCAAGGTCTT 899
QY 1009 gcaaatgttatggtttaaat 1028
Db 900 AAATACATATGGGTGAGAT 919

RESULT 11
AC108131 151941 bp DNA linear PRI 06-FEB-2002
LOCUS Homo sapiens chromosome 16 clone RP11-337N9, complete sequence.
DEFINITION AC108131 AC022860
ACCESSION AC108131.2 GI:18542984
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 6, 2002 this sequence version replaced gi:18369977.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source Location/Qualifiers
1..151941
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-337N9"
BASE COUNT 42554 a 32907 c 31887 g 44593 t
ORIGIN
Query Match 25.2%; Score 281.2; DB 9; Length 151941;
Best Local Similarity 60.4%; Pred. No. 6.6e-62;
Matches 568; Conservative 0; Mismatches 363; Indels 9; Gaps 6;
QY 11 atagagagtggtgaaattggttattattgagaacaatggcagaaaattatgctattctcatta 70
Db 70064 ATTAAATGATAAGGTGGTAATAATTTTGACTATACAATAACTTAACATTGATGATGAAGA 70123
QY 71 gatctctggttttctccacagatat-cctttgctttaattttctctcctaagaatgataa 129
Db 70124 GGTCTATTGCTTTTTCAGGTATATAAAATATAAATTCCTCTTCATCATTAAGACGACCA 70183
QY 130 acttcctggttttattaccatacaaatctcttttaccacaagccagctttggaatttc 189
Db 70184 TGTTCAGTTAAGGTACTACTAACAATTGCCCTTTCTTTTCAAGCTGTAATGGACTCAG 70243
QY 190 agcaaacacccatcctctcttcttccacatcttcacaccttgttttccagtcacaggtctaa 249
Db 70244 AGCCAACAACCTCCTACTATTCTTCCAAATCTTCTCACTCCTTCAGGATCACAGGCCTAA 70303
QY 250 gtccattgacatgataaattagtcacctgtctctcatccacatactgtctgtcttcaatca 309
Db 70304 GCCCACTGACCCCATCACCTGTTCACCTGGCCCTTTGTCACATTAGGGATGTCTCATTGT 70363
QY 310 ggcaaatatggtgctcctagactcttctgttcacagaatactcagatgatcttagta 369
Db 70364 GGCTCTCTTGGCATCTCCACAGCTGTTTGAGTCACTGTATTTTCAGAAATGACTTCAAGTG 70423
QY 370 taaggctcattgtctttttaacaagggtgtagggggcctctccatctgccccctgcct 429
Db 70424 TAAGGCATTCTTCTGTCATGCACAGGGGTGATGAGGAGCTCTCCATCTGCACACCGCT 70483
QY 430 cctgagtgctccagc---catcatcacccacagactctctccttgcacaagctcaa 486
Db 70484 CCTGAGCATGCTCCAGGCTGTGCGCCATCAGCCCTGGCACCTCTCTGGTCTCGAAGATTAA 70543
QY 487 acatcctctgcaagtcacatcttaggattctctcttctctcattgggtgcctcaacatgt- 545
Db 70544 ACAGAAATTCAAAGGTTACATCTTTCACCTCTCTTCTCTCTATGGGTCTCTCAGTTTGTG 70603
QY 546 tcattggtgtaattctgtgtacactgagggtacccccagtgaaacggggccagtcctt 605
Db 70604 TCTCAGTAGTAACCTGCTCTCTCCACTGTGGCTTCTTAATGGGACCAAGAGCGTGT 70663
QY 606 -ctgtttgtcatcacagcactgttcccttttggccatgagctacacccacaggaactgtt 664
Db 70664 GCTAAGTATCAGTAAATACATGATCACTTTCTTCATTAAGTACATATATCAGGAGCCCTGTC 70723
QY 665 ttccacactaatgactttgagggtgctcacctttatagggttccatggtctctcctcaaggg 724
Db 70724 TTTTCATGCTGCCACTGCTTACGAATGTCTTCTTTGTAGCAATCATCATGAGCCCTCAAGTGC 70783
QY 725 ctacatggtgattattttatcacagacaataaagaggttatctcagtgcttcacacagcca 784
Db 70784 ATACATGGTAATTTCTCTTGTTCAGGC-ATCAGAGACAATCCCAATACCTTCATAGAACCA 70842
QY 785 gcctgtccccagctcaccagtgaaaagagcctccccaggtctatcttactgctggtgagt 844
Db 70843 ACCTCTCTCAAGAGCCTCCCCAGAGAAAAGGGGACCA--TCCTGTGCTGGTGAGTTCG 70900
QY 845 ttgtcttcacatactgggtgggactttacgtctctctattttcaggaggtgtgacatgata 904
Db 70901 TTTCTGTCATGTACTGGGGGACCTTATTATCTCATCTCTCTCACTCTATTATTGGCT 70960
QY 905 aatgattctctgctagtgctggctccaggttattgtggcca 944
```


Db 148141 GGCTTCTGGCATCTCCAGACCTGTTTGAGTCACCTGATTTTCAGAAATGACTTCAAGTG 148082
QY 370 taaggtcatgtctcttttaacaaggtgatgaggccctccatctgcacccctccct 429
Db 148081 TAAGGCATCTTCTGCGATGACAGGGTGATGAGAGCCTCTCCATCTGCACACCTGCCT 148022
QY 430 cctgagtgctgcacggc---catcatcagcccccagcatctctccttgccaaagctcaa 486
Db 148021 CCTGAGCATGCTCAGGCTGTCGCCATCAGCCCTGGCACCTCCTGGTCTCGAAGGATTAA 147962
QY 487 acatcctctgcgaagtcaacatcttaggattctctctcttctctcttggtctcacaatgt- 545
Db 147961 ACAGAAATTCAAAAGGTTACATCTTTCACCTCTCTTCTCTCTATGGTTCAGTTTGTG 147902
QY 546 tcattggtgtaattctctctgtacactgcgctacccctaccccccagtgaaacgggcccagctt 605
Db 147901 TCTCAGTAGTAACCTGCTCTCCACCTGTGGCTTCTCTTAATGGGACCAAGAGCGTGT 147842
QY 606 -ctgtttgtcatcagacactgttctcttcttggccatgagctacacacacagagactgtt 664
Db 147841 GCTAAGTATCAGTAATACTGATGATCACTTCTTCATCAAGCTACATTTATCAGGAGCCTGTC 147782
QY 665 tticacactaagacttgaggatgtcacactttatagggttcatggtctctctcaagagg 724
Db 147781 TTTATGCTGCCACTGCTTACGAATGCTTCTTTGTAGCAATCATGCGACCTCAAGTGC 147722
QY 725 ctacatggtgattattttatcacagacaataagaggtctatctcagtgcccttcacacagcca 784
Db 147721 ATACATGTAATCTCTTGTTCAGGC-ATCAGAGACAATCCCAATACCTTCATAGAACCA 147663
QY 785 gctgtcccgagctctcacagtgaaagagctcccccagctctcttactgctggtgagt 844
Db 147662 ACTCTCTCCAGAGCCTCCCGAGAGAAAAGGGACCA--TCTTGTGCTGGTCAAGTGC 147605
QY 845 ttigtcttcacactggtggtgactttacgttctctcatttcagagaggtgacatggata 904
Db 147604 TTCTGCTCATGTACTGCGGGGGACCTTATTCATCTCCTCCTCAACTCTATTATTGGCT 147545
QY 905 aatgattctctgtagtgggtggtccaggttattgtggcca 944
Db 147544 TATGACTCAGTCATGTTGAGCATGCGAGAGGCTTCCTGGCA 147505

RESULT 13
AC022202/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-321E8 from 7, complete sequence.
ACCESSION AC022202
VERSION AC022202.12 GI:13992747
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 95097)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 95097)
AUTHORS Meyer, R., Abbott, A. and Hawkins, M.
TITLE The sequence of Homo sapiens BAC clone RP11-321E8
JOURNAL Unpublished
AUTHORS Waterston, R.H.
REFERENCE 3 (bases 1 to 95097)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 4 (bases 1 to 95097)
AUTHORS Waterston, R.H.
TITLE Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (08-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 95097)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2001 this sequence version replaced gi:13431144.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0321E08

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 17 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR17, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-468B6, 200 bp overlap; the
clone sequenced to the right is RP11-561N12. Actual start of this
clone is at base position 61428 of RP11-468B6; actual end is at
base position 95097 of RP11-321E8.

FEATURES

Location/Qualifiers
1. .95097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="7"
/clone="RP11-321E8"
/clone_lib="RPCI-11"
/rpt_family="Alu"
273. .564
/rpt_family="L2"
1280. .1352
/rpt_family="L2"
1371. .1404
/rpt_family="MER1_type"
1570. .1808
/rpt_family="L2"
1861. .2172

```
repeat_region /rpt_family="Alu"
3218. .3294
/rpt_family="CRI"
repeat_region 3320. .3610
/rpt_family="Alu"
repeat_region 4987. .5293
/rpt_family="Alu"
repeat_region 5727. .6021
/rpt_family="Alu"
repeat_region 6134. .6454
/rpt_family="Alu"
repeat_region 6798. .6945
/rpt_family="MIR"
repeat_region 9181. .9201
/rpt_family="AT_rich"
repeat_region 9239. .9398
/rpt_family="L1"
repeat_region 9472. .9785
/rpt_family="Alu"
repeat_region 9907. .10485
/rpt_family="ERV1"
repeat_region 10594. .10760
/rpt_family="L1"
repeat_region 10936. .11287
/rpt_family="L1"
repeat_region 11288. .11311
/rpt_family="(TAA)n"
repeat_region 11312. .13026
/rpt_family="L1"
repeat_region 13027. .13325
/rpt_family="Alu"
repeat_region 13326. .14023
/rpt_family="L1"
repeat_region 14024. .14328
/rpt_family="Alu"
repeat_region 14329. .14394
/rpt_family="L1"
repeat_region 14395. .14416
/rpt_family="(CAA)n"
repeat_region 14417. .14706
/rpt_family="L1"
repeat_region 14736. .14817
/rpt_family="L1"
repeat_region 14818. .14837
/rpt_family="(GAA)n"
repeat_region 14838. .15485
/rpt_family="L1"
repeat_region 15486. .15770
/rpt_family="Alu"
repeat_region 15771. .16672
/rpt_family="L1"
repeat_region 16662. .16946
/rpt_family="L1"
repeat_region 17098. .17162
/rpt_family="L2"
repeat_region 18516. .18615
/rpt_family="MER1_type"
repeat_region 18609. .18650
/rpt_family="MIR"
repeat_region 18773. .18987
/rpt_family="L1"
repeat_region 19997. .20162
/rpt_family="MER1_type"
repeat_region 20282. .20581
/rpt_family="Alu"
repeat_region 20586. .20743
/rpt_family="L1"
repeat_region 20832. .21571
/rpt_family="L1"
repeat_region 21732. .21798
/rpt_family="L1"
repeat_region 21799. .22104
/rpt_family="Alu"
```

```
repeat_region 22105. .22878
/rpt_family="L1"
repeat_region 22879. .22902
/rpt_family="AT_rich"
repeat_region 22934. .22957
/rpt_family="AT_rich"
repeat_region 23000. .23270
/rpt_family="Alu"
repeat_region 23481. .23523
/rpt_family="AT_rich"
repeat_region 23738. .23769
/rpt_family="AT_rich"
repeat_region 24080. .24144
/rpt_family="L1"
repeat_region 24145. .24451
/rpt_family="Alu"
repeat_region 24452. .25442
/rpt_family="L1"
repeat_region 26510. .26544
/rpt_family="(TTTG)n"
repeat_region 27297. .27572
/rpt_family="L1"
repeat_region 27573. .27870
/rpt_family="Alu"
repeat_region 27871. .28013
/rpt_family="L1"
repeat_region 28014. .28324

Query Match 25.2% Score 280.2; DB 9; Length 95097;
Best Local Similarity 58.8%; Pred. No. 1.1e-61;
Matches 597; Conservative 0; Mismatches 403; Indels 15; Gaps 6;

QY 11 atagagatgttgaaattgggtattattgagaaacatgagcagaataattgctatttcatta 70
DB 36574 ATAAAAATAATCATAAATGGTAATAATTTTGAACACTCACAAGAACATTCATGACACAG 36515
QY 71 gatctctgttttctccacagatctctgttcttaattctcttaagatgatcaaa 130
DB 36514 ATGTATTTTTCAGGTCCTCAAAATATGTTCTTCTACATCATAGTAGTAACACACAT 36455
QY 131 ctctctgtttttattaccatacaaatctcttttaccacagccagcttgggaatttca 190
DB 36454 TTCATATGTTTATCTT---TGAAAAAGGCCCTTTTATTTCAAGCTGGCATTGAATCTCG 36399
QY 191 gcaaacaccatccttcttctccacatcttccacttctgttcttcagtcacaggtctaa 250
DB 36398 GCCAACATATTTCTTCTCTGTCCTGGCACATTTTATATCTTTAAGGATCACAAAGCCTAAA 36339
QY 251 tccattgacatgataatagtcacctgtctctccacatactgctgctcttcaactcag 310
DB 36338 AACCATGACCTGATCATCTGTCACTTGACCTTTGTTTTCATACATGCTAGTCAATTGCA 36279
QY 311 gcaatattggtgtccttagactcttcttgggttcacagaatactcaggatgatcttagtat 370
DB 36278 GCAAAAGCTATTGTCCTCCAGATGGGTTTGGAGTCACAGAATTTTCAGATAACTTCAGATGT 36219
QY 371 aaggtcattgtctttttaaacaagtgatgaggggctctccatctgcaccccccctgcctc 430
DB 36218 AAGGCTGTGTTCTACATACAAAGGCAATGAGGGGCCCTCTTGATCTGCACAAAGCTCTCTC 36159
QY 431 ctgagtgtgtccagggccat---catcagccccagcatcttctcttgggcaagctcaaa 487
DB 36158 CTGAGCATGCTTCAGGGCCATCACCATTAGCCCCAGCACCCTCTGGTGGTGAGATTTAAA 36099
QY 488 catcttctgcagtcacatctttaggattcttcttctcatgggtctcctcaaatgttc 547
DB 36098 CATAAAAATCACAAAATACAAATATCTGCTGCTTTATATCTTTGTCCTCCTCAATTTTCTT 36039
QY 548 attggtgtaattcttctgtgtacactgcggctacccccacagtgaaacggggcagctctct 607
DB 36038 TCAACAGTGACATGATTATCTACATTTGAGGTTTTTCCAGTGTGACCCAG--ATAAATCT 35981
QY 608 gttgttcatacagcactgttctcttttttgcccatgagctacacccacaggaactgtttt 667
```

```

Db 35980 GAATGTGAGTAAATACGTCTCACTTTCCCAATGAATGTCAACATCAGAAAGGCTGTTGT 35921
Qy 668 cacaataagacttgaggatgcacacttatagggttcattggtctctcaagagccta 727
Db 35920 TACTCTCTGCTTATCCAGAGATGCTTCTCTGTAGGAATCATGCCGCTCTCAAGTGCCTA 35861
Qy 728 catggtattattttacagacaataagaggctatctcagtcgcttcacacagccagcc 787
Db 35860 CATGGTGATCTCTGTCCAGGC-ATCAGAGGCGCTCCAGCACCTTCACAGCACTAGCT 35802
Qy 788 tgt--ccccagtgctaccagtgtaaaagagcctccagggctatcttactctggtgagt- 844
Db 35801 TTTTATTAAAGAACCTCCCCAGAGAAAGAGGCGCCACCAAGACCATCTTGCTGCTGGTAGTT 35742
Qy 845 -tttgcttcacatactggtgagactttacgtcttcatttcaggaggtgacatgga 902
Db 35741 TCTTTGGTGTATGACTCATGTGACCTTAATGTCTCATCTCCACAAATGTTGTATGGG 35682
Qy 903 taaatgattctctgctgdtgtggtccaggttatgttggtgccaatagctatgcgcgaatta 962
Db 35681 TATTACGCCCTCATCTACAGTGTCCACAAAGTTTATGGTCAATGCCCTATGCCACTGTCA 35622
Qy 963 gtcccttgatgctaattatgctgataaccaaatattcaagagactctgcaaatgtt 1017
Db 35621 GTCCATGGTGTAAATCAGATCTGATAAAAGAAATCATCAGTATTCTGCCAAAGGT 35567

RESULT 14
AC073188
LOCUS AC073188 65128 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.
ACCESSION AC073188
VERSION AC073188.10 GI:15920117
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 65128)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 65128)
AUTHORS Cordum,H. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-73B2
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 65128)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 65128)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 65128)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:14647308.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0073B02

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mail to: egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

FEATURES	source
Location/Qualifiers	1..65128
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP11-73B2"
	/clone_lib="RPCI-11"
repeat_region	406..1310
	/rpt_family="L1"
repeat_region	1317..1547
	/rpt_family="L1"
repeat_region	3257..3634
	/rpt_family="L1"
repeat_region	3635..3925
	/rpt_family="MaLR"
repeat_region	3930..4120
	/rpt_family="MaLR"
repeat_region	4136..4334
	/rpt_family="Alu"
repeat_region	4341..4488
	/rpt_family="Alu"
repeat_region	4555..4753
	/rpt_family="L1"
repeat_region	5597..5693
	/rpt_family="MERL_type"
repeat_region	5802..6205


```
repeat_region /rpt_family="L1"
9188..9281
/rpt_family="MER53"
10528..10758
/note="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
repeat_region /rpt_family="MaLR"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="(TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/note="similar to EST AI806860 (NID:g5393426) wf36b07.x1"
repeat_region /rpt_family="AT_rich"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="(TTTTTA)n"
14129..14883
/rpt_family="L1"
15223..15246
/rpt_family="AT_rich"
15290..15571
/note="similar to EST AW445087 (NID:g6986849)"
misc_feature /note="similar to EST AW627961 (NID:g7374751) hi36d12.x1"
16305..16574
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="(CAAAA)n"
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="(A)n"
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="(GAAA)n"
19256..19691
/rpt_family="ERV1"
19748..19790
/rpt_family="(CAAAA)n"
20557..20613
/note="similar to EST AW445087 (NID:g6986849)"
misc_feature /note="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
20594..20828
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/note="similar to EST BF899282 (NID:g12290741)"
22640..23050
/note="similar to EST AI279713 (NID:g3917947) ql53f03.x1"
```

```
misc_feature 22855..23086
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
repeat_region /rpt_family="T-rich"
23402..23451
repeat_region /rpt_family="Alu"
23430..23733
23850..24148
/rpt_family="MER2_type"
24566..25150
/rpt_family="L1"
25151..25275

Query Match 25.1%; Score 279.6; DB 9; Length 65128;
Best Local Similarity 62.6%; Pred. No. 1.6e-61;
Matches 553; Conservative 0; Mismatches 319; Indels 12; Gaps 7;

Qy 140 ttattaccatacaaatctctttttatccacaagccagctttggaatttcagcaaacacc 199
Db 12154 TGTATGTTATTTAAAGAGCTGCTTTGTTTACAGTGTGTTGGAATCTCAGTCAATACT 12213
Qy 200 atccttctttttccacatcttccacctgttttttcagtcacaggtctcaagtcattgac 259
Db 12214 TTTCTGTCTCTTTCAGCATCTTCCACTCTCTCTGTATCACAGACTAAACCCACACTGAC 12273
Qy 260 atgataatagtcacctgtctctccacatactgtctcttcacactcagcgcaatttg 319
Db 12274 TTGATCATCTGTCCTTGCCTCTGTCACATAGTAACCTTCTTCATTGCACTGTATTG 12333
Qy 320 gtgccttagactcttttgggttcacagaatactcagatgatcttaggtataaagtcatt 379
Db 12334 CAGTCTCCAAATCTGTTGTCATCATTTGAA-CTTCAGAAATGACTTCAAGTGTAAAGCACTT 12392
Qy 380 gtcttttaacaagggtgatgggggctctccatcttcacacctgcacccctgcctcctcagtg 439
Db 12393 TTCCACCTGAGCAGGGTGTGTGGGGCTCTCTATCTGCACACACTGCCCTCAGTGTG 12452
Qy 440 ctccagg---ccatcatcagcccgagctctctctctggcaagctcaaacatcctct 496
Db 12453 TTCAGGCCACCAACATCAGCCCCAGCACCTCCCTGATGGCAAGGTTTAAACATAAATCT 12512
Qy 497 gcaagtcacatcttaggattctctctctttctcctgggttcctcacaacatggttgcgt-gt 555
Db 12513 TCAAAATTACATTATCCATGTTTCTTCTTCTTATGTTGTTCTTCAACTTGTCTTTCAGTAGT 12572
Qy 556 aatctctgtgtacactggtgtacccccagtg-aaacgggggagctctctctgttgc 614
Db 12573 AACAGGATCTTCTACACTGTAGTCTTCCAGTGTGAGCCAGACAAATGTACTGAATATA 12632
Qy 615 atacgacactgtctcttttggcccatgagctacacccacagagagactgttttcacacta 674
Db 12633 GGTAAAGTACTGCTCAGTGTCTTCTCCATGAGCTTCACTCAGGGGCCACGTTTCTTATTCGT 12692
Qy 675 atgactttgagggtatgcacctttatagggttcatgggttcctctcagaagggctacatggtg 734
Db 12693 ACAACACTTAGGGATGCTTTCTTGTAGGAATACTGCTGTTCTTAATTGCATACATGCTG 12752
Qy 735 attattttatacagaataaaggctctctcagtcgcttcacacagccagccctgtcccc 794
Db 12753 ACTCTCTATTTCAGGC-CTCAGAGGTGATCCACACCTTCCACAGCAATGGTCTCTTCCCC 12811
Qy 795 --gagctcacagtgaaagagctccaggctctctcttactctctggtgagttt---tgt 849
Db 12812 AGGAGTTCCTCCAGAAAACGGGCCATCCAGACCACTACTGCTGTGGTAAAGTTTCTCTGT 12871
Qy 850 cttcacactagtggtgaccttaacgttctctcttttcagggaggtgtgacatggataaaatga 909
Db 12872 GGTCTCTGTGTGGGTGATTTTCATCATCTCTCTCTCCTCAGCAATGTTGTGGCATATGG 12931
Qy 910 ttctctgtctag-tgtggtccaggttatgtggccaatagctatgccgcaattagtccttt 969
Db 12932 CCCAATCAACCAAGGTTACAGAGTCTTGTAGTCAATGTCTATGCCGTGATCAGTCTCT 12991
Qy 970 gatgctaatttatgctgataaccacaaatattcaagactctgcaaa 1013
```


Db 15886 ATTAAATGATAAGGTTGGTAATAATTTTGACTACACAAATAACTAACATTTGATGATAAGA 15945
Qy 71 gatctctgtctttttctccacagatat-cctttgtctttaattttctcttaagatgatcaa 129
Db 15946 GGTCTATTGCTTTTTCAGGTATATAAAATATAATCTCTTCATCATTAAGACGACCA 16005
Qy 130 acttctgtggtttattaccatacaaaatctctttttatccacaagccagcttttggaaatttc 189
Db 16006 TGTTCACAGTTAAGGTACTACTAACAAATTCGCCCTTCTTTTCAAGCTGGTATTTGGACTCAG 16065
Qy 190 agcaaacaccatctcttcttttccacatcttccacctttgtttttcagtcacagagctctaa 249
Db 16066 AGCCAACAACATTCCTCCATTCTTCCAAATCTCTCACTCTCTCAGGATCACAGGCCTAA 16125
Qy 250 gtccattgacatgataattagtcacctgtctctctcatccacataactgctgctcttcaactca 309
Db 16126 GCCCACTGACCCCATCACTCTGACCTGGCCCTTGTCCACTTAGGGATGCTCCTCATTTGT 16185
Qy 310 ggcaatatgtgtctcttagactcttttgggttcacagaataactcaggatgactcttaggta 369
Db 16186 GGTCTTCTTGGCACTCCAGACCTGTTTGGTCACTGTATTTTTCAGAAATGACTTCAAGTG 16245
Qy 370 taaggctcatgtcttttttaaacaagtgatgaggggctctccatctgcacccccctgct 429
Db 16246 TAAGGCATTTCTTGCATGCACAGGGTGATGAGGAGCCTCTCCATCTGCACCACCTGCCT 16305
Qy 430 cctgagtgctccaggc---catcatagccccagcatctctctcttggcacaagctcaa 486
Db 16306 CCTGAGCATGCTCCAGGCTGTGGCATCAGGCCCTGGCACCTCCTGGTCTGCAAGGATTAA 16365
Qy 487 acatctctctgcaagtcacatcttaggattctcttcttcttctcattggttcccaacatgt- 545
Db 16366 ACAGAAATTCAAAGGTTACATCTTTCACTCCTCTCTCTCTATGCGGTCTCAGTTTGTG 16425
Qy 546 tcattggtgtaactctctgtctgtacactgctgctacccccagtgaaacggggccagctctt 605
Db 16426 TCTCAGTAGTAACCTGCTCTCTCCACTGTGGCTCTTCTTAATGGGACCAAGACGGTTGT 16485
Qy 606 -ctgttgcatacagacactgtctcttttgcacatgagctcacacccacagagactgtt 664
Db 16486 GCTAAGTATACAGTAAGTAATACTGATCACTTTCTTCCATAAGCTTACATTTATCAGGAGCCTGTC 16545
Qy 665 ttccacacataatgactcttgagggtgtcaccttttataggggttcattggtcctctcaagagg 724
Db 16546 TTTTCATGCTGCCACTGCTTACGAATGCTTCTTTGTAGCAATCATGACAGCCCTCAAGTGC 16605
Qy 725 ctacatggtgattattttatcacagacaataagaggctatctcagtgcccttcacacagcca 784
Db 16606 ATACATGGTAATTTCTTTGTTTCAGGC-ATCAGAGACAATCCCAATACCTTTCATAGACCA 16664
Qy 785 gctgtccccagagtcacacagtgtaaaagagcctccacaggctatcttactgtggtgagt 844
Db 16665 ACCTCTCTCCAAGAGCCTCCCCAGAGAAAAGGGGACCA--TCTTGTGCTGCTGAGTTGC 16722
Qy 845 ttgtcttcacataggttgaggttactcttctcatttccagaggttgacatgata 904
Db 16723 TTTCTGGTCATGTACTGGGGGGAGCTTATTTATCTCATCTCTCTCAACTCTATTATTGGCT 16782
Qy 905 aatgattctctgtagtggtgctccagggttattgtggcca 944
Db 16783 TATGACTCAGTCATTGTGAGCATGCAGAGGGCTTCTGGGCA 16822

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 01:51:05 ; Search time 107.09 Seconds
(without alignments)
2555.194 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	62.4	5.6	7218	1	US-08-232-463-14
C 2	56.2	5.0	8353	3	Sequence 14, Appli
C 3	48.4	4.3	246240	2	Sequence 1, Appli
C 4	48.4	4.3	246240	2	Sequence 20, Appli
C 5	48.4	4.3	246240	2	Sequence 21, Appli
C 6	47.4	4.2	638	4	Sequence 22, Appli
C 7	44.4	4.0	497	4	Sequence 275, App
C 8	43.3	3.9	152331	3	Sequence 109, App
C 9	40.6	3.6	2270	2	Sequence 16, Appli
C 10	40.6	3.6	13674	2	Sequence 1, Appli
C 11	40.6	3.6	17410	1	Sequence 3, Appli
C 12	40.6	3.6	17410	1	Sequence 3, Appli
C 13	40.6	3.6	17410	1	Sequence 3, Appli
C 14	40.6	3.6	17410	2	Sequence 3, Appli
C 15	40.6	3.6	17410	2	Sequence 3, Appli
C 16	40.6	3.6	17415	3	Sequence 1, Appli
C 17	40.6	3.6	17415	5	Sequence 1, Appli
C 18	39.3	3.5	504	4	Sequence 120, App
C 19	36.2	3.2	80246	4	Sequence 4, Appli
C 20	36.2	3.2	80595	4	Sequence 3, Appli
C 21	35.4	3.2	289	4	Sequence 17, Appli
C 22	35.4	3.2	289	4	Sequence 17, Appli
C 23	35.2	3.2	2358	3	Sequence 10, Appli
C 24	35.2	3.2	2360	4	Sequence 15, Appli
C 25	35.2	3.2	40352	3	Sequence 9, Appli
C 26	34.2	3.1	4002	3	Sequence 29, Appli
C 27	33.6	3.0	437	4	Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptz9pt-F1s
; IMMEDIATE SOURCE:
US-08-232-463-14

Sequence 29, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 7, Appli
Sequence 6, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli

```
Query Match          5.6%; Score 62.4; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. NO. 1.7e-08;
Matches 21; Conservative 228; Mismatches 159; Indels 0; Gaps 0;

Qy 149 atacaaatctctttatccacagcagcttggaaatttcagcaaacacacatctctt 208
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1030 ATTCGAGCTGGCTGAGGTCGAGGAGCTGCGATYYYYYYYYYYYYYYYYYY 1089

Qy 209 cttttcacatcttcacatcttggtttcagtcacaggtctcaagtcattgacatgataatt 268
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1090 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1149

Qy 269 agtcacatgtctctcaccacataactgctgctctcactcaggaataattggtgcctta 328
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1150 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1209

Qy 329 gactcttgggtcacagaatactcagatgatttaggtataagtcattgtctttta 388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1210 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1269

Qy 389 acaagtgtaggggcctctcactcgcacccctgcctcctgagtgctccagcc 448
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1270 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1329

Qy 449 atcacagccacagcatctctctcttgcaagctcaaacatcctctcgaagtcacatc 508
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1330 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1389

Qy 509 taggattctctcttctcagtggtcctcaacatgcttcattggtga 556
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1390 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTGTA 1437

RESULT 2
US-08-611-587-1/c
; Sequence 1, Application US/08611587
; Patent No. 6150091
; GENERAL INFORMATION:
; APPLICANT: PANDOLFO, MASSIMO
; APPLICANT: MONTERMINI, LAURA
; APPLICANT: MOLTO, MARIA D.
; APPLICANT: Koenig, Michael
; APPLICANT: Campuzano, Victoria
; APPLICANT: Cossee, Mireille
; TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,587
; FILING DATE: 03-MAR-1996
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Brashears-Macatee, Sarah J.
; REGISTRATION NUMBER: 38,087
; REFERENCE/DOCKET NUMBER: D-5901
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5620
; TELEFAX: 713-651-5246
; TELEX: 76-2829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8353 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 9q13
; UNITS: bp
; US-08-611-587-1

Query Match          5.0%; Score 56.2; DB 3; Length 8353;
Best Local Similarity 83.1%; Pred. No. 1.3e-06;
Matches 64; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1013 atgttatggtttaaatattgtctctcctccaaagctcattgttgaaatttaatcgccaatgt 1072
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7323 ATGCTATGTTGTTGAATGTTGTACCTCCGAAACTCATGTTGAAACTTAATTCCTCCGATGT 7264

Qy 1073 ggcagttactaagaagt 1089
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7263 GGCAGTATTGAGAGGTG 7247

RESULT 3
US-08-724-394A-20/c
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
```

[illegible]

Sequence 16, Application US/08852807
Patent No. 5861298
GENERAL INFORMATION:
APPLICANT: Debouck, Christine
APPLICANT: Drake, Fred
APPLICANT: Gowen, Maxine
APPLICANT: Rood, Julie
APPLICANT: Hastings, Gregg
APPLICANT: Adams, Mark
APPLICANT: Fraser, Claire
APPLICANT: Lee, No. 5861298man
APPLICANT: Kirkness, Ewen
APPLICANT: Blake, Judith
APPLICANT: Fitzgerald, Lisa
TITLE OF INVENTION: CATHEPSIN K GENE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,807
FILING DATE: 07-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,942
FILING DATE: 14-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/020,273
FILING DATE: 17-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/026,273
FILING DATE: 26-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50006-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2270 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-852-807-16

Query Match 3.6%; Score 40.6; DB 2; Length 2270;
Best Local Similarity 77.8%; Pred. No. 0.027;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggttttaaatattgtctctccaaagctcatgttgaatttaataatgcgaatgtgg 1074

Db 576 GATATGTTGGTGTGTGTGTCCTCCCTCCAGTCATCTGAAATGTGATCCCAATGTGG 517

Qy 1075 cag 1077

||

Db 516 GAG 514
RESULT 10
US-08-852-807-1/c
Sequence 1, Application US/08852807
Patent No. 5861298
GENERAL INFORMATION:
APPLICANT: Debouck, Christine
APPLICANT: Drake, Fred
APPLICANT: Gowen, Maxine
APPLICANT: Rood, Julie
APPLICANT: Hastings, Gregg
APPLICANT: Adams, Mark
APPLICANT: Fraser, Claire
APPLICANT: Lee, No. 5861298man
APPLICANT: Kirkness, Ewen
APPLICANT: Blake, Judith
APPLICANT: Fitzgerald, Lisa
TITLE OF INVENTION: CATHEPSIN K GENE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,807
FILING DATE: 07-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,942
FILING DATE: 14-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/020,273
FILING DATE: 17-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/026,273
FILING DATE: 26-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50006-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 13674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-852-807-1

Query Match 3.6%; Score 40.6; DB 2; Length 13674;
Best Local Similarity 77.8%; Pred. No. 0.07;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggttttaaatattgtctctccaaagctcatgttgaatttaataatgcgaatgtgg 1074


```

;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
;
; US-07-841-646-3
;
; Query Match          3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.079;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; Qy 1030 ttgtctctcctcaagctcatgttgaaatttaacgcgaatgtggcagctactaagaagtg 1089
;      ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGTCCCTACAAAACACTGTTGAAATTTAAATGCCAGTGTAAACATTAATTGAGAGCTT 8871
;
; Qy 1090 atg 1092
;      |||
; Db 8872 atg 8874
;
; RESULT 12
; US-08-147-023-3
; Sequence 3, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
```

```

; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-147-023-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.88; Pred. No. 0.079;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; Qy 1030 ttgtctctccaaagtcattgttgaaatttaacgcaatgtggcagtgactaagaagtg 1089
; ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGCTCCTACAAAACACTGTTGAAATTAAATGCGCAGTGTAACATTATTGAGAGGTT 8871
;
; Qy 1090 atg 1092
; |||
; Db 8872 ATG 8874
;
; RESULT 13
; US-08-447-570-3
; Sequence 3, Application US/08447570
; Patent No. 5714589
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,570
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
;
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
```

NAME/KEY: misc_feature
LOCATION: 10960..10961
OTHER INFORMATION: /label= GAP-2
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11059..11211
OTHER INFORMATION: /label= EXON-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-447-570-3
Query Match 3.6%; Score 40.6; DB 1; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.079;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1030 ttgtctctccaaagctcatgttaatttaacgcgaatgtgcagtaactaagaagtg 1089
||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTTGTCCCTACAAAACATCATGTTGAAATTTAATGCCAGTGTACATTATTGAGAGCTT 8871
QY 1090 atg 1092
|||
Db 8872 ATG 8874
RESULT 14
US-08-449-700-3
Sequence 3, Application US/08449700
Patent No. 5863758
GENERAL INFORMATION:
APPLICANT: OPPERMANN, HERMANN
APPLICANT: OZKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS

COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,700
FILING DATE: 21-FEB-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 810,560
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 827,052
FILING DATE: 28-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 660,162
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,988
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,849
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 616,374
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 600,024
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 599,543
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 579,865
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 569,920
FILING DATE: 20-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 483,913
FILING DATE: 22-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 422,613
FILING DATE: 17-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 315,342
FILING DATE: 23-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 232,630
FILING DATE: 15-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 179,460
FILING DATE: 08-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-001CP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
FEATURE:

```
NAME/KEY: exon
LOCATION: 3192..3730
OTHER INFORMATION: /label= EXON-1
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10413..10414
OTHER INFORMATION: /label= GAP-1
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
THIS SEQUENCE."
NAME/KEY: exon
LOCATION: 10696..10891
OTHER INFORMATION: /label= EXON-2
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10960..10961
OTHER INFORMATION: /label= GAP-2
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
THIS SEQUENCE."
NAME/KEY: exon
LOCATION: 11059..11211
OTHER INFORMATION: /label= EXON-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
THIS SEQUENCE."
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
THIS SEQUENCE."
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-449-700-3
```

```
Query Match 3.68; Score 40.6; DB 2; Length 17410;
Best Local Similarity 77.8; Pred. No. 0.079;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Caps 0;
Qy 1030 ttgtctctccaaagctatgttgaaatttaacgcgaagtgtggaagtactaagaagt 1089
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8812 TTGTCCTACAAAACATCATGTGAATTTAATGCGCACTGTAACATTTATTCAGAGTT 8871
Qy 1090 atg 1092
|||
Db - 8872 ATG 8874
RESULT 15
```

```
US-08-449-699A-3
; Sequence 3, Application.US/08449699A
; Patent No. 5958441
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,699A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/147,023
; FILING DATE: 01-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: STK-001CP6CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-449-699A-3
```

```
Query Match 3.6%; Score 40.6; DB 2; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.079;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1030 ttgtctctccaaagctcatgttgaaatttaatcgccaatgtggcagtgactaagaagtg 1089
||||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTTGTCCCTACAAAACACTCATGTTGAAATTTAATGCGCAGTGTAACATTATTGAGAGGTT 8871
||||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1090 atg 1092
|||
Db 8872 ATG 8874
```

Search completed: July 21, 2002, 03:56:34
Job time: 7529 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 00:21:30 ; Search time 4107.22 Seconds
(without alignments)
5675.898 Million cell updates/sec

Title: US-09-728-309-1

Perfect score: 1114

Sequence: 1 attccagatcatagagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: gb_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

RESULT 1

AL390728

LOCUS

DEFINITION

PROGRESS ***

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

ALIGNMENTS

AL390728 206832 bp DNA linear HTG 06-FEB-2002
Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
PROGRESS ***

AL390728 206832 bp DNA linear HTG 06-FEB-2002

HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (sites)

Ashwell, R.

Direct Submission

Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 7, 2002 this sequence version replaced gi:18491329.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA488L18
 ----- Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Chemistry: Dye-terminator Er-amersham; 9% of reads Chemistry: 100% of reads
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 206638; 4.3% error; agarose-fp
 Quality coverage: 9.91x in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.
 Location/Qualifiers
 1. 206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-488L18"
 /clone_lib="RPC1-11.2"
 1. 204689
 /note="assembly_fragment:03301"
 204790. 206832
 /note="assembly_fragment:03801"
 BASE COUNT 58135 a 44577 c 44388 g 59631 t 101 others
 ORIGIN
 misc_feature
 misc_feature
 Query Match 98.4%; Score 1096.4; DB 2; Length 206832;
 Best Local Similarity 99.6%; Pred. No. 1.8e-273;
 Matches 1110; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
 QY 1 attccagatcatagagatgttgaaattggtattattgagaaacatggcagaaattatgct 60
 Db 144680 ATTCCAGATCATAGAGATGTTGAAATTGGTTATTATTGAGAACATGGCAGAAATTATGCT 144739
 QY 61 attctcattagatctctgtctttctccacagatataccttgccttaattttcctctaa 120
 Db 144740 ATTCTCATTAGATCTCTTGGCTTTCTCCACAGATATCCTTTGCTTTAATTTCCTCTCAA 144799
 QY 121 gatgatcaaaccttccgtggtttttattaccatacaaatctcttttattccacagcagctt 180
 Db 144800 GATGATCAAACTCTCTGGTTTATTACCAATCAAAATCTCTTTTATCCAAAGCCAGCTT 144859
 QY 181 tggatttcagcaaacaccatcctctcttttccacatcttcaccttggtttcagtcga 240
 Db 144860 TGGAAATTCAGAAACACCATCTCTCTTTTCCACATCTTCACCTTTGTTTCAGTCA 144919
 QY 241 cagcttaagtccattgacatgataatagtcacctgtctctctccacatactgtgct 300
 Db 144920 CAGGCTAAGTCCATTGACATGATAATAGTCACTCTCTCTCATCCACATACCTGCTCT 144979
 QY 301 ctccactcaggcaaatatggtgctcttagactctcttggttcacagataactcaggatga 360
 Db 144980 CTTCACATCAGCAATATGGTGCTCTTAGACTCTCTTTGGTTTCAGAAATACTCAGGATGA 145039
 QY 361 tcttaggataaaggtcattgtctcttttaacaagggtgatggggcctctccatctgcac 420
 Db 145040 TCTTAGGTATAAGGTCTATTGCTCTTTTAAACAAGGTGATGAGGGGCTCTCCCATCTGCAC 145099

QY 421 cccctgctcctgagtgctgtccaggccatcatcagccccagcatcttctcttggcaaa 480
 Db 145100 CCCCTGCTCTGAGTGTGCTCCAGGCGCATCATCAGCCCCAGCATCTTCTCTTGGCAAA 145159
 QY 481 gctcaaacatcctctctgcaagtcacatcttaggattcttctcttctcattgggtcctcaa 540
 Db 145160 GCTCAAAACATCCTTCTGCAAGTCACATCTTAGGATTCTTCTCTTCTCATGGGTCTCAA 145219
 QY 541 catgttctcattggtgtaattctctgctgtacactgcggctacccccagtgaaacggggccca 600
 Db 145220 CATGTTTCATGGTGTAACTCTCTGTACACTGCGGCTACCCCGAGTGAACGGGGCCCA 145279
 QY 601 gtctctgtttgtccatcacagcactgttctctttttgccccagtcagtcacccccagggagac 660
 Db 145280 GTCTTCTGTTGTACACAGCACTGTTCTTTTGGCCCATGAGCTACACCCACAGGAGAC 145339
 QY 661 tgttttcacactaatgactttgaggatgtcacttctcctttatagggttcattggtctctctcaa 720
 Db 145340 TGTGTTTTCACACTAATGACTTTTCAGGGATGTCACCTTTATAGGGTTCATGGTCTCTCAA 145399
 QY 721 gaggtacatggtgatatttttatcacagacaataaagggtctatctcagtcgctctcagca 780
 Db 145400 GAGGCTACATGGTGATTAATTTATACAGACAATAAGAGGCTATCTCAGTGCCTTCACACA 145459
 QY 781 gccagcctgtccccgagctctcaccagtgaagagagctcccgaggtctattctctgtggt 840
 Db 145460 GCCAGCCTGTCCCCGAGTCTCACCAAGTGAAGAGAGCTCCCGAGGTATCTTACTGCTGT 145519
 QY 841 gagtttcttcacatactggtggtgactttacgttctctcatttctcagggaggtgacatg 900
 Db 145520 GAGTTTGTCTTTCACATACACTGGGTGACTTTACGTTCTCATTTTTCAGGAGGTGTGACATG 145579
 QY 901 gataaatgattctctgctagtgctgctcaggttatttggccaaatagctatgcgcgaat 960
 Db 145580 GATAAATGATTCTCTGCTAGTGTGCTCCAGGTATTGTGGCAATAGTATGCCGCAAT 145639
 QY 961 tagtctcttgatgctaaatttatgctgataaccataattattcaagactctgcgaatgttatg 1020
 Db 145640 TAGTCTTTGATGCTAATTTATGCTGTGAACCAATATTTCAAGAGCTCTGCAATGTATG 145699
 QY 1021 gtttaaatatttctctcctccaaagctcatgttgtaatttaacccaatgtggcagttac 1080
 Db 145700 GTTTAAATATTGTCTCTCCAAAGCTCATGTGTAATTTAATCCCAATGTGGCAGTAC 145759
 QY 1081 taagaagtgatgatgagagggttaattccattcatg 1114
 Db 145760 TAAGAAG---TGATGAGAGGTTAATCCATTCTATG 145790

RESULT 2

AL359674/c
 ID AL359674 standard; DNA; HTG; 166798 BP.
 AC AL359674;
 SV AL359674.4
 XX
 DT 20-JUN-2000 (Rel. 64, Created)
 DT 20-JAN-2001 (Rel. 66, Last updated, Version 5)
 XX
 DE Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-433K2
 DE
 KW HTG; HTGS_DRAFT; HTGS_PHASE1.
 XX
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 XX
 RN [1]
 RA Plumb B.;
 RT
 RL Submitted (20-JAN-2001) to the EMBL/GenBank/DBJ databases.

FT		/note="assembly_fragment:00356"	
FT	misc_feature	148878..151529	
FT		/note="assembly_fragment:00391"	
FT	misc_feature	151630..155541	
FT		/note="assembly_fragment:00444"	
FT	misc_feature	155642..159186	
FT		/note="assembly_fragment:01041"	
FT	misc_feature	159287..162766	
FT		/note="assembly_fragment:01610"	
FT	misc_feature	162867..166798	
FT		/note="assembly_fragment:01755"	
XX			
SQ	Sequence 166798 BP; 45603 A; 37045 C; 37398 G; 43934 T; 2818 other;		
Query Match 98.1%; Score 1093.2; DB 30; Length 166798;			
Best Local Similarity 99.5%; Pred. No. 1.2e-272;			
Matches 1108; Conservative 0; Mismatches 3; Indels 3; Gaps 1;			
OY	1	attccagatcatgagatgttgaaattggttattattgagaacatgccagaaaattatgct 60	
Db	154134	ATTCCAGATCATGAGATGTGAAATGGTTATTATTGAGAACATGCCAGAAAATTATGCT 154075	
OY	61	atttcattagatccttggtttttctcacagatatcctttggttaatttcctttaaa 120	
Db	154074	ATTTCATTAGATCTCTTGTTTTCTCCACAGATATCCTTGCCTTAATTTCCTCTTAA 154015	
OY	121	gatgatcaaaacttcctggtttttattaccatacaaaattcttttatccacaagcagctt 180	
Db	154014	GATGATCAAACTTCCTGGTTTTATTACCATAAATAATCTTCTTTATCCAACAGCGACTT 153955	
OY	181	tggaattcaagcaaacacocatcctttctttccacatcttcacaccttggtttttcagtca 240	
Db	153954	TGGAATTTACAGCAACACCATCCTTCTCTTTCCACATCTTCACCTTTGTTTTTCAGTCA 153895	
OY	241	caggcttaagtccaatgcagatgaataattagtcacctgtctctccatccacatactgctgt 300	
Db	153894	CAGGCTTAAGTCCATTGCATGATGAATAATTAGTCACCTGTCTCTCATCCACATACTGCTGCT 153835	
OY	301	cttcactcaggcaaatattgggtcccttagactctttgggttcacagaatactcaggatga 360	
Db	153834	CYTCCCCAGGCANATATGGTGCTTAGACTCTTTGGTTTCACAGAATACTCAGGATGA 153775	
OY	361	tettagggtataagggtcattgtctcttttaaacaagggtgatgaggggctctccatctgtcac 420	
Db	153774	TCTTAGGTATAAGGTCAATGTCTTTTAAACAAGGTGATGAGGGGCTCTCCATCTGCAC 153715	
OY	421	ccccgtcctcctgagtggtcctccaggccatcatcagccccagcatcttctccttggcaca 480	
Db	153714	CCCCGTCCCCTTGAGTGTGCTCCAGGCCATCATPACGCCCCAGCATCTCTCTCTGGCAA 153655	
OY	481	gctcaaacatcctctcgaaagtcaecatcttaggattctccttttctcatgggtcctcaa 540	
Db	153654	GCTCAACATCCTCTGCAAGTCACATCTTAGGATTTCTTCTTTCTCATGGGTCTCTCAA 153595	
OY	541	catgttcattggtgtaattctctgtgttacactgcggctacccccccagtgaaacggggcca 600	
Db	153594	CATGTTCAATTGGTGTAAATCTTCTGTGTACACTGCGGCTACCCCCAGTGAAACGGGCCA 153535	
OY	601	gtctctgttgtcatacagaactgttctcttttggccatgagctacaccccagcagagac 660	
Db	153534	GTCCTTCTGTTGTACAGCACTGTTCCTTTTGGCCCATGAGTACACCCACAGGAGAC 153475	
OY	661	tgttttccacctaagtacttggaggatggtcacacctttatagggttcattgggtcctctcaa 720	
Db	153474	TGTTTTTCACACTATGACTTTGAGGGATGTCACCTTTATAGGGTTCTATGGTCTCTCAA 153415	
OY	721	gaggctacatggtgattattttatcacagacaataagaggctatctcagtgcttccacgca 780	
Db	153414	GAGGCTACATGGTGATTTATTTTATACAGACAATAAGAGGCTATCTCAGTGCCTTCACACA 153355	
OY	781	gccagcctgtccccgaggtctcacccagtgaaagacctccccagggtatatcttactgctggt 840	

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4074: contig of 4074 bp in length
 * 4075 4174: gap of unknown length
 * 4175 9708: contig of 5534 bp in length
 * 9709 9808: gap of unknown length
 * 9809 13642: contig of 3834 bp in length
 * 13643 13742: gap of unknown length
 * 13743 52846: contig of 39104 bp in length
 * 52847 52946: gap of unknown length
 * 52947 140430: contig of 87484 bp in length
 * 140431 140530: gap of unknown length
 * 140531 217087: contig of 76557 bp in length.

FEATURES

source

1. .217087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-433K2"
 /note="assembly_name:Contig168"
 1. .4074
 /note="assembly_name:Contig169"
 4175. .9708
 /note="assembly_name:Contig169"
 9809. .13642
 /note="assembly_name:Contig170"
 13743. .52846
 /note="assembly_name:Contig171"
 52947. .140430
 /note="assembly_name:Contig172"
 140531. .217087
 /note="assembly_name:Contig173"

BASE COUNT 60435 a 48981 c 48285 g 58724 t 662 others
 ORIGIN

Query Match 98.18; Score:1093.2; DB 2; Length 217087;
 Best Local Similarity 99.58; Pred. No. 1.2e-272;
 Matches 1108; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
 Qy 1 attccagatcagagatgttgaattggtattattgagacatgagcagaaattatgct 60
 Db 121154 ATTCAGATCATAGAGATGTTGAATTGGTATTATTAGACACATGGCAGAAATATGCT 121095
 Qy 61 attctcattagatctctgtctttctccacagatataccttggcttaattttcctcttaa 120
 Db 121094 ATTCATATTAGATCTTGTCTTTCTCCACAGATATCCTTTGCTTTAATTTTCTTCTTAA 121035
 Qy 121 gatgataaacttctcgtttttattaccatacaaaatcttctttttaccagaagccagctt 180
 Db 121034 GATGATCAAACTTCTGCTTTTATTACCATATAAATCTCTTTTATCCACAAGCCAGCTT 120975
 Qy 181 tggaaattcagcaaacacatctctcttcttccacatcttccaccttggttttcagtcga 240
 Db 120974 TGGAAATTTCCGAACACACATCCTTCTCTTTTCCACATCTTCACCTTGTGTTTCAGTCA 120915
 Qy 241 caggctcaagtcattgacatgataattagtcacactgtctctctcaccacatactcgtcgt 300
 Db 120914 CAGGCTCAAGTCCATGATGATATAATTAGTCACCTGTCTCTCATCCACATACCTGCTGCT 120855
 Qy 301 ctteactcaggcaaatattggtccttagacttcttgggttcacagaatactcaggatga 360
 Db 120854 CTTTACCACGCAATATTGGTGTCTTGTAGACTTCTTTGTGTTTACAGAAATACTCAGGATGA 120795
 Qy 361 tcttaggtataaggtcattgtcttttttaaaacaggtgatgaggggacctctccatctgcac 420
 Db 120794 TCTTAGGTATAAGGTCAATGCTTTTAAACAGAGTGATGAGGGGCTCTCTCCATCTGCAC 120735
 Qy 421 cccctgcctcctgagtgctccaggccatcatcagccccagcagcatcttctccttggcaaa 480
 Db 120734 CCCCTGCCTCCTGAGTGTCTCCAGGCCATCATCAGCCCCAGCATCTTCTCTCTTGGCAAA 120675

Qy 481 gctcaaacatcttctgcaagtcacatcttaggattcttcttcttctcattgggtctctcaa 540
 Db 120674 GCTCAAAATCATCTTCTGCAAGTCACATCTTAGGATPCTTCCTTTTCTCATGGGTCTCTCAA 120615
 Qy 541 catgttcattggtgttaattctctgctgtacactgctggtacaccccgagtgaaacggggcca 600
 Db 120614 CATGTTCAATGGTGAATCTTCTGCTGTACACTGGGGCTACCCCGAGTGAAACGGGGCCA 120555
 Qy 601 gtctctctgtttgttcatacagcaactgttctctttttgtcccatgagctacacccacagagac 660
 Db 120554 GTCTTCTCTTGTTCATACAGCACTGTCTCTTTTGTGCCATGAGCTACACCCACAGAGAC 120495
 Qy 661 tgtttttcacactaatgactttgagggatgacaccccttattaggggttcattggtccctctcaa 720
 Db 120494 TGTFTTTTTCACACTAATGACTTTTGGAGGATGTCACCTTTATAGGGTTCATGGTCTCTCAA 120435
 Qy 721 gaggtacatggtgattattttatcacagacaataagaggtcatctcagtgcccttcacgca 780
 Db 120434 GAGGCTACATGGTGATTATTTATACACACAATAGAGGCTATCTCAGTGCTTCACACA 120375
 Qy 781 gccagctgtcccccaggtctcaccagtgtaaaagagcctccaggctattcttactgctggt 840
 Db 120374 GCCAGCCTGTCCCCGAGTCTCACCAGTGAAGAGAGCCTCCAGAGGCTATCTTACTGCTGT 120315
 Qy 841 gagttttgttcacatactggtggtgactttacgtttctcattttcaggaggtgtgacatg 900
 Db 120314 GAGTTTGTCTTTCACATACACTGCGGTGGACTTTTACGTCTCATTTTCAGAGGTTGACATG 120255
 Qy 901 gataaatgatctctcgtcagtggtccaggttatgtggccaatagctatgcccgaat 960
 Db 120254 GATAAATGATCTCTGCTAGTGTGGCTCCAGGTTATTTGGCCATAGCTATGCGCGCAAT 120195
 Qy 961 tagtctttgatgctaatattatgctgataccaaatattccaagactctgcaaatgttatg 1020
 Db 120194 TAGTCTTTGATGCTTAATTTATGCTGTATAACCAAAATATTCAAGACTCTGCAAAATGTTATG 120135
 Qy 1021 gtttaaatattgtctctccaaagctcattgttgaatttaattcccaatgagcagctac 1080
 Db 120134 GTTTAAATATTGTCTCTCTCCAAAGCTCATGTTGAATTTAATGCCAATGTGGCAGTAC 120075
 Qy 1081 taagaagtgatgatgagaggttaattccattcatg 1114
 Db 120074 TAAGAAG--TGATCAGAGGTTAATCCATTTCATG 120044

RESULT 4

AL390728/c
 LOCUS AL390728 206832 bp DNA linear HTG 06-FEB-2002
 DEFINITION Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.
 ACCESSION AL390728
 VERSION AL390728.22 GI:18614040
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)
 Direct Submission
 Ashwell,R.
 Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Feb 7, 2002 this sequence version replaced gi:18491329.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA488L18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Et-amersham; 9% of reads Chemistry:
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.9ix in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.

FEATURES

source
 1..206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-488L18"
 /clone_lib="RPC1-11.2"
 1..204689
 /note="assembly_fragment:03301"
 204790..206832
 /note="assembly_fragment:03801"
 58135 a 44577 c 44388 g 59631 t 101 others

Query Match 48.6%; Score 541; DB 2; Length 206832;
 Best Local Similarity 77.7%; Pred. No. 2e-129;
 Matches 813; Conservative 0; Mismatches 205; Indels 29; Gaps 12;
 QY 1 attccagatcatagaga----tctgaaattggtttattattgagaacatggcagaaa--- 53
 Db 126599 ATTCCAGATCAGACAGATTATTGACATATATGATCTAATAATTAACACATGCGCAGAAATGT 126540
 QY 54 ----tatgctatctcattagatctctctgtctttctccacagatatcctcttcttaatt 110
 Db 126539 TCCTAATGCAATTATCATTTGATGCTTATTATTTCTCTACAGATATCTTTATTTGTTT 126480
 QY 111 ttccttctaag--atgatcaaatctcctggtttattaccacataatctctttttatc 167
 Db 126479 AATTTTCCATGGAATGACCAAGCTTTTCAGCTATATTGTCTATTAATAATGTTTATTACC 126420
 QY 168 cacaagccagcttttgaatttcagaaacacatctctctctcttttccacatcttcaact 227
 Db 126419 CTCAGTCAGCTTTGGATCTCAGCAACACCTTCCTCTCTTTTCCACATCTTCACCT 126360
 QY 228 ttgttttcagtcacaggtctaatgctcattgacatgatattagtcacctgtctctcacc 287
 Db 126359 TTGCTTATACACAGGCTTAAGCCATTGACATGACCATTTAGTCACCTTGCCCTTAATCC 126300
 QY 288 acatactgctctcttactcaggaataattggtgtccttagactcttcttggttcacaga 347
 Db 126299 ACATACGTCTACTCTTCACTCAGCGGACTACTGCTGCTCCGAGCTATTGTAATCATGGA 126240
 QY 348 atactcag---gatgctcttaggtataggtcattgtctcttttttaacaaggtgatggg 404
 Db 126239 ATATTCAGACAATGATCTAAGTGTGAAGATCATACATTTTAAACAGGGTGATGAGG 126180
 QY 405 gctctccatctgcacccctgctcctgagtggtgtctccagggccat---catcagcccca 461
 Db 126179 GAGTCTCCATCTGCACCACTTGCTCTCTAAGCGTGCTCCAGGCCATCACCATCAGCCCCA 126120

QY 462 gcattctctcttggaagctcaaacatctcttgcgaagtcacatcttaggtattctcc 521
 Db 126119 GCACCTCTCTTCTGGAAGAAATTTAAACATATTTCTGCAATACACACCTTAGCTTCATCC 126060
 QY 522 tttctcatgggtctcacaatgttcattggtg-taatcttctctgtacatcggcgcta 580
 Db 126059 TCFTCTCATGGGTCTCAACATGTTCAATTAATACCTTTTCTTTCATTGTGCTTA 126000
 QY 581 cccccagt-gaaaagggggccagctctcttctgtctacacagcactgttctcttttggcca 639
 Db 125999 CCCGGAATAGATTGGGGCCAGTCTTTTGTGTCACCTGAGCACTGTTATGTTTGGCCA 125940
 QY 640 tgaactacacccacagagagactgttttttcacataaacttgaggatgtcaccctta 699
 Db 125939 TGAGCTACACCCACAGAGGCTGTTTTCATATAATAGTGTGTTTGGGGATGTCATCTCA 125880
 QY 700 taggggttcattgctctcacaagaggtcacatggtgattattttatcacagacaataaagg 759
 Db 125879 TAGGACTCATGGTCCCTCAAGTGGCTA-TGGGTGATATTATTTGTACAAACA-AAGAGA 125822
 QY 760 ctatctcagtccttcacagcagcagcgtctcccc--gagctccaccagtgaaaagagcc 817
 Db 125821 CTGTCCAGCACCTACATAGCACCGCTCTCCCCAAGAGCCTCCCCAGCAAAAAGGGCC 125762
 QY 818 tccaggtctattctactgctggtga---gtttgtcttcacactcgggtggactttacg 874
 Db 125761 TCCAGACCATCTATTTGTTGGTGGAGCTGTTTGTCTTATATATTTGGTGGACTTCATA 125702
 QY 875 ttctcattttcagaggtgtgacatgataaattgattctctgctagtgctcaggtt 934
 Db 125701 TTCTCATTTTCAGTGGTGTGACATGATGAACGCCCTCTACTGTGATGATTCATATG 125642
 QY 935 attgtggccaatagctatccgcaattagtcctttgacttaatttctgctgataaccac 994
 Db 125641 CCTGTGTCATAAATATGTCATATTTAGTCTTTAGCTTTACCTATGCTGATAA-CAA 125583
 QY 995 atattcagaactctcgaatgttatgg 1021
 Db 125582 ATATTCAAGATTGTTCAAGACTCTTTGG 125556

RESULT 5

AC104335
 LOCUS Homo sapiens chromosome 1 clone RP11-433K2, WORKING DRAFT SEQUENCE, 217087 bp DNA linear
 DEFINITION AC104335
 ACCESSION AC104335 AL359674
 VERSION AC104335.1 GI:17426299
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 217087)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 217087)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (08-DEC-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 COMMENT On Dec 8, 2001 this sequence version replaced gi:9801069.

 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: http://www.genome.washington.edu
 Contact: uwgchgs@u.washington.edu
 Drafting Center: SC

 Project Information
 Center project name: chr-1
 Center clone name: RP11-433K2 (sc0158)

 Summary Statistics

Sequencing vector: plasmid; 50% of reads
Chemistry: Dye-terminator ET; 64% of reads
Chemistry: Dye-terminator Big Dye; 36% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 21380 bases at least Q40
Consensus quality: 215084 bases at least Q30
Consensus quality: 216045 bases at least Q20
Insert size: 256066; 17.2% error; agarose-fp
Insert size: 216587; sum-of-contigs
Quality coverage: 6.3x in Q20 bases; agarose-fp
Quality coverage: 7.5x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

- 1 4074: contig of 4074 bp in length
- * 4075 4174: gap of unknown length
- * 4175 9708: contig of 5534 bp in length
- * 9709 9808: gap of unknown length
- * 9809 13642: contig of 3834 bp in length
- * 13643 13742: gap of unknown length
- * 13743 52846: contig of 39104 bp in length
- * 52847 52846: gap of unknown length
- * 52947 140430: contig of 87484 bp in length
- * 140431 140530: gap of unknown length
- * 140531 217087: contig of 76557 bp in length.

FEATURES

source

1. .217087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPCI human BAC library 11"
/clone_id="RP11-433K2"
1. .4074
/note="assembly_name:Contig168"
4175. .9708
/note="assembly_name:Contig169"
9809. .13642
/note="assembly_name:Contig170"
13743. .52846
/note="assembly_name:Contig171"
52947. .140430
/note="assembly_name:Contig172"
140531. .217087
/note="assembly_name:Contig173"
60435 a 48981 c 48285 g 58724 t 662 others

BASE COUNT
ORIGIN

Query Match 33.7%; Score 375.6; DB 2; Length 217087;
Best Local Similarity 77.9%; Pred. No. 1.6e-86;
Matches 547; Conservative 0; Mismatches 134; Indels 21; Gaps 7;

QY 1 attccagatcatagaga----tgttgaattggttattattgagaacatggcagaaaa--- 53
Db 139231 ATCCAGATCACAGATATTTCACATAATGATATACTAATAACACATGGCAGAAATGT 139290

QY 54 ---ttatgtattcttagatctcttcttcttccacagatatactttgttttaatt 110
Db 139291 TCCTAATGCAATATCATTTGATGCTTATTATTTCTCTACAGATATCCTTTATTTGTTTT 139350

QY 111 ttctcttaag---atgatcaaatctccgtttttattaccatacaaatctttttatc 167
Db 139351 AATTTCCATGGAAATGACCAAGCTTTTCAGCTATATTGTCTAATAAATGTTTATTACC 139410

QY 168 cacaagccagcttggatttcagcaaacaccatctcttcttccacatcttcacct 227

Db 139411 CTCAAGTCAGCTTTGGAATCTCAGCAACACCTTCCTCCTCTCTTTTCCACATCTTCACIT 139470
QY 228 ttgttttcagtcacaggttcttaagtccattgacatgaataattagtcacctgtctctcatcc 287
Db 139471 TTGCTTATATCTCAGAGCTTAAGCCCATTGACATGACCATAGTACCTTAGTCTGCCCCATATCC 139530
QY 288 acatactgtgtctcttcactcaggcaaatattggtgtcccttagaactctcttggttcacaga 347
Db 139531 ACATACTGTCTACTCTTTCACTCAGGCGATATCTGGTGTCTCCCGACTTTATTTTGAATATGGA 139590
QY 348 atactcag---gatgatcttaggtataaggtcattgtcttttttaaaacaggtgtgatgaggg 404
Db 139591 ATATTGAGAACAAATGATCTAAAGTGTAAGATCATCATACATTTTAAACAGGGTGATGAGGG 139650
QY 405 gctctccatctgcacccctgctcctctgagtggtgtcctcagggcat---catcagcccca 461
Db 139651 GAGTCTCCATCTGCACCACTTGCTCTCTAAGCGTGTCTCCAGGCGATCAGCATCAGCCCCA 139710
QY 462 gcatctctctctggcacaagctcaaacatctctctcaagtcacatcttaggattcttcc 521
Db 139711 GCACCTCTCTTTGGAAAAGTTTAAACATATTCTTCAATACACCTTAGGCTTCATCC 139770
QY 522 tttctcatggtgtccatcatgttctcattggt-gtaattctctgtgtacactgcggcta 580
Db 139771 TCTTCTCATGGGTCTTCAACATGTCTCATTAATAACCTTTTGTCTTTTCATTTGCTTA 139830
QY 581 cccccagt-gaaacggggccagtcctctgtttgtctacacagcactgtctctttttgccca 639
Db 139831 CCCCAGAATAGGATTTGGGGCAGTCTTTTGTCTCTCAGCAGCTGTATGTGTTTGGCCA 139890
QY 640 tgaagtcacacccacagagacgttttttcacactaatgactt 681
Db 139891 TGAGCTACACCCACAGGAGCCTGTTTTTCATACTAAGGCCTT 139932

RESULT 6
AC074050/c 73661 bp DNA linear PRI 04-JAN-2002
LOCUS Homo sapiens chromosome 16 clone CTD-2358C21, complete sequence.
DEFINITION AC074050
ACCESSION AC074050
VERSION AC074050.3 GI:18057079
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 73661)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 73661)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 73661)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jan 4, 2002 this sequence version replaced gi:15963644.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality


```
Query Match      28.0%; Score 311.6; DB 9; Length 152748;
Best Local Similarity 62.7%; Pred. NO. 6e-70;
Matches 638; Conservative 0; Mismatches 359; Indels 21; Gaps 9;

QY 15 agatattgaattgattattgagacatgacgagaattatgcttattctcattagatc 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132487 AAAGATGATGATGGAACAAATTTGAACACTGACACCAACATGATGAACATGGATG 132546

QY 75 tcttgctt-----tctccacagatatcttctgctttaaatttcttccctcagaatgatca 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132547 TATGTTCTTTTATTCAGATATAGAAGATATGTTCTCTCTATGCCATTAAGATGAAC 132606

QY 129 actcttggttttattaccatacaaaattcttcttaccagaagccagctttggaattt 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132607 TATTTTCTACTTATGCTATCGTTAAATAATGCGCTTGTGTTTTCGAAGCTGCATTTGGATTCT 132666

QY 189 cagcaaaacaccatctctcttctccacatcttcaccttcttcttcttctcagtcacaggtcta 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132667 CAGCCNACACCTTCTCTCTTCTTCTCCACATCTTTTACGGTCTCTTCTGAATCCGAGCCCTTA 132726

QY 249 agtccattgacatgaattagtcacctgtctctcaccacacatactgctgctctcactc 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132727 AACCCCGTGACTTACTCACCCTGCTCACCCTGCCTCATTCACATTCAGATGCTCCTCACATG 132786

QY 309 aggcaaatattggtctcttagactctcttcttcttcttctcagaataactcaggatgatcttaggt 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132787 CAGTGGATTTTGTGCTCTAGACATATTTGAATCATCTGCAATTTTGGGAATGACTTTTAAGT 132846

QY 369 ataaggtcatctcttctttaaagagtgatgagggcctctccatctcaccctcctcgc 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132847 GTAAGGCACCTTTTACACAAACAGGGCAATGAGGGCTCTCCATCTCACACACCTGCC 132906

QY 429 tcttgagtgctccaggccat---catcagcccccagcatctctctctcttcttcttcttctc 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132907 TTCTGAACATGCTCCAGGCGCATCAGCATCAGCCGCCAGCACCTCTCTGTTGGCAAGTTTA 132966

QY 486 aacatctctcagagtcacatcttaggattctctcttctcttctcattcagggctcctcaacatgt 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132967 ACATATAATCCAAATTAATTTGTCATGTTTCTCTTTTGGGTGCTCTCAATTTGT 133026

QY 546 tcaatctgtctgtacactgcggctaccoccccggtgaaacgggg---ccagtc 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133027 CCTTCAGTAGTTGGCCAACTTTTCTCTGTCGATCTTCCCAACAGACCCATATCAACG 133086

QY 604 tcttggttgcacagcactgttctcttcttgcacatgagctacacccacagagagactgt 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133087 TACAGAAATGTCAGTAAATACTGCCACTTTC-CCCCATGAATCCCAATCAGGGGAGTGT 133145

QY 664 tttcacactaatgactttgagggatgtcaccttttatagggtttcattgctctcctcaagag 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133146 TTTTCTACTGTGACATTATCCAAAGTATGCTCTCTGTGTAGTACTTATGCTGTCTCAAGGG 133205

QY 724 gctacatggtgattttttatcacagacaataagagctatctcagtgcccttcacgcagcc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133206 CATACATGTTGATTTCTCTGCCAGCG-ATTGGAGGCAATCCAGACCTCTCATAGCACCC 133264

QY 784 agcctgtcccc--gagctcaccagtgaaagagcctccagggctatcttactgtctgtgtg 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133265 CACCTCTCCCAAGACCCCTCCCAAGAAAGGGCCACCCAGACCATCTGCTGTGTTG 133324

QY 842 agt---tttgctctcacactcaggttgagcttttagcttctcattttcaggaggtgtgaca 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133325 AGTTCTTTTGTGGTGATGATCTGTTGGGTGGAGCTTATCTTCTCAACCTCTTTTACCCCTGCTTA 133384

QY 899 tggataaatgattctc---tgcctagtggtgctccaggttatttcttggtgccaatagctatgcc 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133385 TGGCCGTATGACCCACACGTTTCTCTGAGTGTGTCAGAGAGGCTGTGCTCAATGCCATATGCC 133444

QY 956 gcaattagctctttagtgtaatttatgtcgataacccaataattcgaagactctgcataa 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133445 ACTGTTAGTCTCTTTGTGGCAAAATCAGTTCTGTGATAAAACAATAATCATGATATTATGCAAA 133502
```

```
RESULT 8
AC073188/c
LOCUS
DEFINITION
AC073188 Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.
AC073188 65128 bp DNA linear PRI 09-JAN-2002
VERSION
AC073188.10 GI:15920117
KEYWORDS
HTG..
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 65128)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE
2 (bases 1 to 65128)
Cordum,H. and Abbott,A.
The sequence of Homo sapiens BAC clone RP11-73B2
Unpublished (2001)
REFERENCE
3 (bases 1 to 65128)
Waterston,R.H.
Direct Submission
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 65128)
Waterston,R.H.
Direct Submission
Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
5 (bases 1 to 65128)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gl:14647308.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0073B02
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/UTB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

```
source
1. 65128
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-73B2"
406..1310
/rpt_family="L1"
1317..1547
/rpt_family="L1"
3257..3634
/rpt_family="L1"
3635..3925
/rpt_family="MaLR"
3930..4120
/rpt_family="MaLR"
4136..4334
/rpt_family="Alu"
4341..4488
/rpt_family="Alu"
4555..4753
/rpt_family="L1"
5597..5693
/rpt_family="MER1_type"
5802..6205
/rpt_family="L1"
9188..9281
/rpt_family="MER53"
10528..10758
/note="similar to EST AI028332 (NID:g3245641) ov90903.xl"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11645..11696
/rpt_family="TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/note="similar to EST AI806860 (NID:g5393426) wf36b07.xl"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="(TTTTTA)n"
14129..14883
/rpt_family="L1"
```

```
repeat_region 15223..15246
rpt_family="AT_rich"
misc_feature 15290..15571
note="similar to EST AW445087 (NID:g6986849)"
misc_feature 15290..15571
note="similar to EST AW627961 (NID:g7374751) hi36d12.xl"
repeat_region 16305..16574
rpt_family="Alu"
repeat_region 16753..17021
rpt_family="Alu"
repeat_region 17038..17286
rpt_family="Alu"
repeat_region 17700..18467
rpt_family="L1"
repeat_region 17853..17872
rpt_family="(CAAAA)n"
repeat_region 18468..18778
rpt_family="Alu"
repeat_region 18752..18796
rpt_family="(A)n"
repeat_region 18779..18938
rpt_family="L1"
repeat_region 18838..18870
rpt_family="AT_rich"
repeat_region 18944..19246
rpt_family="Alu"
repeat_region 19220..19265
rpt_family="(GAAA)n"
repeat_region 19256..19691
rpt_family="ERV1"
repeat_region 19748..19790
rpt_family="(CAAAA)n"
misc_feature 20557..20613
note="similar to EST AW445087 (NID:g6986849)"
misc_feature 20557..20666
note="similar to EST AI028332 (NID:g3245641) ov90903.xl"
misc_feature 20594..20828
note="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
repeat_region 21125..21428
rpt_family="Alu"
repeat_region 21664..21691
rpt_family="AT_rich"
misc_feature 22442..22750
note="similar to EST BF899282 (NID:g12290741)"
misc_feature 22640..23050
note="similar to EST AI279713 (NID:g3917947) ql53f03.xl"
misc_feature 22855..23086
note="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
repeat_region 23402..23451
rpt_family="T-rich"
repeat_region 23430..23733
rpt_family="Alu"
repeat_region 23850..24148
rpt_family="MER2_type"
repeat_region 24565..25150
rpt_family="L1"
repeat_region 25151..25275
```

Query Match 27.2%; Score 303; DB 9; Length 65128;
Best Local Similarity 62.0%; Pred. No. 9.5e-68;
Matches 638; Conservative 0; Mismatches 365; Indels 26; Gaps 9;

```
QY 11 ataggagattgaaattggtattattgagaacatggcagaattatgtattcttcatta 70
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37223 ATTAAATGATAATATGTTGTAATAATTTGACCGTGAATAATAATAATATGATGATAGGA 37164

QY 71 gattctgtcttttccacagatatcccttgccttaattttccttcttaagaatgatcaaa 130
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37163 TATCATTTGCTTTTTCAGATATATAAAATATGATCCCTCTTCCACCATTAAGACGCCATA 37104

QY 131 cttcctgggttttattaccatacaaatcttctttttatcccaagccagctgttggaaattca 190
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37103 G-TCCACTTATGACACTATTAAATAATGCCCCCTTTTCCCAAGCTGGTATTGGACTACA 37045
```


QY 371 aaggctcattgtctttttaaacbaagtgatgagggccctctccatctgcaacccccctgcctc 430
DB 118279 AAGGTATTCTTACATGCATGAGGGATGAGGG--CTCTCCATCTGCACCACTGCCTC 118336
QY 431 ctgagtgctgcctcaggc---catcataagcccccagcatctctctcttggaagctcaaa 487
DB 118337 CGAGTGTGCTCCAGCGTGTCCACCATATACCCCGGACGCTCTGTTGGCAAGAAATTAA 118396
QY 488 catctcttgcaagtcacatcttaggattctctctctttctctctggtctcacaatgttc 547
DB 118397 CAGAAATTCACACATTCGATCTTTCACCTTCTTGGGTTCTCAGTTT-----CTTTC 118447
QY 548 attggtgtaattctctgtgtacatctggtgtaccccccaagtgaaacggg-gccagttctc 606
DB 118448 TCAGTAGTAACCTGCTTCCTCCACTGTGGCTCTTCTTAATGTGACCGACACAAATGTGC 118507
QY 607 tttgtgtacacagcactgtctcttttttggccatgactgtacacccacaggaagactgttt 666
DB 118508 CAAGTATCAGTAATACTGTTTCACCTTCTTACATAAGCTACATCATCAGGAGTCTGTCT 118567
QY 667 teacataatgacttggagggatgcatcttctttagtgggtctctctctcaagaggtc 726
DB 118568 TCATGCTGCCACTACTTGCAGATGCTCTTCTTGTAGCAGTATGTGCTCTCAAGGGCAT 118627
QY 727 acatggtgattattatatacacacaataagagctatctcagtgctctcagcagccagc 786
DB 118628 ACATGGTAATCTTTTGTGCCAGC-ATCAGACACAACCCGACGTACTTCCCGACCAAC 118686
QY 787 ctgtgc---cccgagtctcaccagtgaaaagagcctccagggctattctactgctggtgagt 844
DB 118687 CTGTCTGCAAAAGCTCCCGCAGAGAAAGGGCCACCTAGACCATTCCGATGCTGTGTGAGT 118746
QY 845 ----tttgtcttcacatactgggtgactttacgtgtctatcttcttcaggaggtgtgacatgg 901
DB 118747 TCGTTTGTGCTCATATATGGGTGGGCTCTCATCTCATCTCTCAACCCCTGTATTGG 118806
QY 902 atcaatgattctctgtagtggtggtccaggttatgtggccaatagctatccgcaatt 961
DB 118807 GCTTATGACTCATTCATCATGCGGTGTCAGATGCTTATGGGCAATTTCTAT---CTGTT 118862
QY 962 agtctttgatgctaattatgctgataccaaaattatcaagactctgcaaatgttatgg 1021
DB 118863 AGCCCTTTTATGCTAATAGTCTGTATAGTATAGATAAATAATCAAAATTTTGCAATATGTTA 118922
QY 1022 ttttaatat 1030
DB 118923 TATAAATAT 118931

RESULT 10
LOCUS AB064662 989 bp DNA linear MAM 05-JAN-2002
DEFINITION Capra hircus gene for putative pheromone receptor gVIR1, complete cds.
ACCESSION AB064662
VERSION AB064662.1 GI:18148935
KEYWORDS
SOURCE Capra hircus (strain:Shiba) DNA.
ORGANISM Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Caprinae; Capra.
1 (sites)
Wakabayashi, Y., Mori, Y., Ichikawa, M., Yazaki, K. and Hagino-Yamagishi, K.
Gene of goat putative pheromone receptor is expressed in two distinct olfactory organs
Unpublished
2 (bases 1 to 989)
Wakabayashi, Y., Mori, Y., Ichikawa, M., Yazaki, K. and Hagino-Yamagishi, K.
Direct Submission

JOURNAL Submitted (03-JUL-2001) Kimiko Hagino-Yamagishi, Tokyo Metropolitan Institute of Medical Science, Department of Ultrastructural Research; 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan (E-mail:k.yamagishi@insoken.or.jp, Tel:81-3-3823-2101(ex.5345), Fax:81-3-3823-2965)
FEATURES
source Location/Qualifiers
1..989
/organism="Capra hircus"
/strain="Shiba"
/db_xref="taxon:9925"
1..927
/gene="gVIR1"
1..927
/gene="gVIR1"
/note="Shiba goat putative pheromone receptor gVIR1"
/codon_start=1
/product="putative pheromone receptor gVIR1"
/protein_id="BAB83523.1"
/db_xref="GI:18148935"
/translation="MKKNRLSSSIDIRNAIFSEVAIGILANTILLFHAHNPFLLEHRP
KSLDTTGHLLALIHIVMLLTAFMATDFTGSKTWDDIQCKLVYLYLSSMGLSICAT
CLLSVLAITLSPNSCLAKFKLRPSHNLYCTFLFWFNPFNGSFLSTIATPNVT
SAHLRLTESCSLRPIVHFRLYLOFALFTIQDILCLGLMALSSGYMTLLYLRHKRQTQ
ELQSTRSPKASPERATQTILFLMSFFVMYFLDITVSWFSRMLWDIDSVRMCVQML
VGNGYASISPLVLISLTKRIIKVLKYIWRWH"
BASE COUNT 240 a 267 c 193 g 289 t
ORIGIN
Query Match 27.1%; Score 302; DB 4; Length 989;
Best Local Similarity 63.2%; Pred. No. 1.2e-67;
Matches 581; Conservative 0; Mismatches 325; Indels 14; Gaps 7;
QY 119 aagatgatacaactctcctggtttttattaccatacaaaattctttttatccacaagccagc 178
DB 4 AAAAAAATAGACTTCCAGTTCCATTGACATGAAGAAATGCCATTTTCTCCGAAGTTGC 63
QY 179 ttgtgaatttcagcaaacaccatctcttcttccacatcttccactttgttttcaagt 238
DB 64 ATTGGATCTTGGCCCAACACCATCTCTTCTCTCCATGCCCAACACACTCTCTTTGAG 123
QY 239 cacaggtctaaatccattgacatgataattagctacactgtctctccatccacatactgctg 298
DB 124 CACAGACCCCAAGTCCACTGACCTGACCAATTTGGTCACTTGGCCCTTAATCCACATAGTCATG 183
QY 299 ctcttcactcaggcaaatattggtgtccttagactcttcttgggttcacagaataactcagat 358
DB 184 CTACTAACTGTGGC---ATTTATGGCTACAGACACCTTTGGGTCTCAGAAAAACCTGGGAT 240
QY 359 gatcttaggtataaggtcattgtctttttaacaaggtgatgaggggctctccactctgc 418
DB 241 GACATCCCAATGTAAAGTTGGTTGTCTACCTGTACAGTTTCATGAGGGGCTCTCCATTGT 300
QY 419 acccctgctcctctgagtggtctccaggccatca---tcagcccccagcatctctctcttg 475
DB 301 GCCACTGCTGCTGAGTGTCTCTGGGCCCATCACCTTCAGCCCCAGAAACTCTCTGTCG 360
QY 476 gcaagctcaaacatctctctgcaagtcacatttaggattcttctcttctcattcaggttc 535
DB 361 GCCAAGTTCAAACTTCGCCCATCCCATCAACAATCTGTATTGCACATCTCTCTCTGGGTC 420
QY 536 ctcaacatgttctcagtggtgtaatacttc-tgctgtacactgcggtaccacccagtg-aaac 593
DB 421 TTCAACATGTTTCATCAATGGTTCCCTTCTTACTGTGCACATTTGCCACCCCAACGACGACC 480
QY 594 gggggcagctctctgtttgtctacacagcaactgtctcttttgcctcagctgagctacacccc 653
DB 481 TCAGCCCATCTCTACGTGTCACTGAATCTCTCACTTCGCCCCCGCTGATTCACTTCCTT 540
QY 654 aggagactgtttttcacactaatgacttttgaggatgtcacactttataggggttcaggttc 713
DB 541 AGGTATTTACAGTTTGCACACTGAGGACCATTCGAAGACATCTGCTTGTGTGGGCTCATGGCC 600

```
Qy 714 ctctcaagaggctacatggtgattattttatcacagacaataaagggtctatctcagtgctt 773
Db 601 CTCTCAAGTGTGATACATGCTGACTCTCTTGTACAGGC-ATAAGAGGCAGAGCCAGCATCT 659
Qy 774 tcacgcagcagcgtgtccccc--gagtcacacagtgtaaaagagcctccacaggctatctt 831
Db 660 TCAGAGCACCAGGCTTCTCCAAAAGCATCCCAAGAAAGAGGGCCACCCAGACCATCTT 719
Qy 832 actgctggtgagt---tttgtcttcacatactgggtgacctttacgttctcattttcagg 888
Db 720 ATTCTCATGAGTTCCTTGTGGTCATGTACTTTTGGACATAACTGTCTCTTGGTTCTC 779
Qy 889 aggtgtgacctggataaattattctctgctagtggtgctccaggttattgtggccaatag 948
Db 780 AAGAATGTTGGGATATCGACTCTGTTGATGTGTGTCAGATGCTTGTGGGAAATGG 839
Qy 949 ctatgcgcgaattagtccttttgatgctaaattattgctgataacaaatattcaagactct 1008
Db 840 CTATGCCAGCATAGCCCTTTGGTGTAAATCAGTACTGAAAGAGAAATAATCAAGGTCTT 899
Qy 1009 gcaaatgttatggtttaat 1028
Db 900 AAAATACATATGGTGAGAT 919

RESULT 11
AC108131
LOCUS Homo sapiens chromosome 16 clone RP11-337N9, linear PRI 06-FEB-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-337N9, complete sequence.
ACCESSION AC108131
VERSION AC108131.2
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 6, 2002 this sequence version replaced gi:18369977.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
Location/Qualifiers
1. .151941
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-337N9"
BASE COUNT 42554 a 32907 c 31887 g 44593 t
ORIGIN
```

```
Query Match 25.1%; Score 279.6; DB 9; Length 151941;-
Best Local Similarity 60.3%; Pred. No. 1.2e-61;
Matches 567; Conservative 0; Mismatches 364; Indels 9; Gaps 6;

Qy 11 atagagagtggtgaaatggttattattgagaacatggcagaataattatgctattctcatta 70
Db 70064 ATTAAAAATGAAGGTGGTAATAATTTTGACTATACAATAAATAACATTAACATTCATGATAAGA 70123

Qy 71 gatctctgctttctccacagatat-cctttgcttttaattttctctcttaagatgataa 129
Db 70124 GGTCTATTGCTTTTTCAGGTATATAAAATATAAATTCCTCTTCATCATTAAGACGACCA 70183

Qy 130 acttcctggtttattaccatacaaatctctttttatccacaagcagcttttggaatttc 189
Db 70184 TGTTCAGTTAAGGTACTACTAAACAATTGCCCTTCCTTTTCAAGCTGCTATTGGACTCAG 70243

Qy 190 agcaaaacacatcctctctttttccacatcttcacacatttcacacagtcacaggtctaa 249
Db 70244 AGCAACAACCTCTCTCTCTATCTTCTTCCAAATCTTCTCACTCTCTTCAGGATCACAGGCCATA 70303

Qy 250 gtccattgacatgataattagtcacacctgtctctcaccacacatactctgctctctcactca 309
Db 70304 GCCCACTGACCCCATCACCTGTCACTGGCCCTTGCCACTTAGGATGCTCTCTCATGT 70363

Qy 310 ggcaattggtgctcctggaactcttttggttcacagaaatactcaggtatgattaggtta 369
Db 70364 GGTCTTCTTGGCATCTCCAGACCTGTTGAGTCACCTGATTTTTCAGAAATGACTTCAAGTG 70423

Qy 370 taaggtcattgctttttaaacaagtgatgagggccctctccatctcactgacccccctgcct 429
Db 70424 TAAGGATCTTCTGTGCATGCACAGGGTGATGAGGAGCCCTCTCCATCTGCACCCAGCTCT 70483

Qy 430 cctgagtggtctccaggc---catcatcagccccagcactctctctctcttgcaaaagctcaa 486
Db 70484 CCTGAGCATGCTCCAGGCTGTGCGCCATAGCCCTGCGCACCTCTGCTCTGCAAGGATTA 70543

Qy 487 acatcctctgcaagtacacatctttaggtattctctctctctctctctctcactcaactgt- 545
Db 70544 ACAGAAATTCAAAGGTATACATCTTTCACCTCTTCTTCTCTCTATGGTCTCTCAGTTTGT 70603

Qy 546 tcattggtgtaattctctgctgacactcggtcagctcccccagtgaaacggggccagcttt 605
Db 70604 TCTCAGTAGTACCTGCTCTCTCTCCACTGTGGCTTCTTCTAATGGGACCAAGCGTTGT 70663

Qy 606 -ctgtttgtcacagcactgttctctttttggccatgagctacacccacagagactgtt 664
Db 70664 GCTAAGTAGTACATAAATACTGATCATCTTCTTCCATAAGCTACATTATCAGAGGCTGT 70723

Qy 665 ttccacataatgaacttgagggtgacacatttatagggttcattggtctctcctcaagagg 724
Db 70724 TTTATGCTGCCACATGCTTACGAATGTCTCTTTGTAGCAATCATGCAGCCCTCAAGTGC 70783

Qy 725 ctacatggtgattattttatcacagacaataaaggctctatctcagtgcttcacgcagcca 784
Db 70784 ATACATGCTAATCTCTTGTTCAGGC-ATCAGAGACAATCCCAATACCTTTCATAGAACA 70842

Qy 785 gctgtccccagtgctcaccagtgaaaagagcctccagggtctattactctgctggtgagt 844
Db 70843 ACCTCTCTCCAAGAGCCCTCCAGAGAAAAGGGGACCA--TCTTGTCTGCTGGTGAAGTGC 70900

Qy 845 ttgtcttcacatactgggtgagactttacgttctctcatttttcagagaggtgtgacatggata 904
Db 70901 TTTCTGGTCAATGTAAGTGGGGGACCTTATTATCATCTCTCTCAACTCTATTATTGGCT 70960

Qy 905 aatgattctctgctagtggtgctccaggttatttgagcca 944
```



```
Db 148141 GGTCTCTTGCCATCCAGACCTGTTTGAGTCACTGATTTTCAGAACTGACTCAAGTG 148082
Qy 370 taaggtcattgtctcttttaaaacagggtgatggggccctctccatctgcaccccccctgcct 429
Db 148081 TAAGGCATTCCTCTGCATGCACAGGGTGATGAGGAGCCTCTCCATCTGCACCACTGCCT 148022
Qy 430 cctgagtgctccaggc---catcatcagcccccagcatctctcccttgcaagagctcaa 486
Db 148021 CCTGAGATGCTCCAGGCTGCGGCATCAGCCCTGGCCACCTCTCTGGTCTGCTCAAGAGATTAA 147962
Qy 487 acatcctcttgcaagtcacatcttaggattctctctctctctctctctctctctctcaacatgt- 545
Db 147961 ACAGAAATTCAGAGTTACATCTTTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147902
Qy 546 tcaatggtgtaattctctctgctgacactgagggctaccccccaagtgaaacggggccagctctt 605
Db 147901 TCTCAGTAGTAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147842
Qy 606 -ctgtttgtctacagcactgttctctctctctctctctctctctctctctctctctctctctct 664
Db 147841 GCTAAGTATCATGTAATACTGATCATCTTCTTCTCATTAAGCTACATTTATCAGGACCTGTC 147782
Qy 665 ttccacactaatgactttgagggtatgacactcttataggggtctatggtctctctctctcaaggg 724
Db 147781 TTTGATGCTGCACTGCTTACGAATGTCTCTTTGTAGCAATCATGCAGCCCTCAAGTGC 147722
Qy 725 ctacatggtgattatttatcacacataagaaggtctctctctctctctctctctctctctctctca 784
Db 147721 ATACATGGTAATCTCTTGTTCAGGC-ATCAGAGACAAATCCCAATACCTTCATAGAACCA 147663
Qy 785 gcctgtcccccaggtctcaccagtgtaaaagagcctcccgaggtctctctctctctctctctctg 844
Db 147662 ACCTCTCTCCAAAGCCTCCCCACAGAAAGGGGACCA--TCTTGTGCTGGTCAAGTGC 147605
Qy 845 ttgtcttccactactcaggggtgactttacgtctctctctctctctctctctctctctctctgata 904
Db 147604 TTTCTGTGTCATGTCATCTGGGGGACCTTATTATCTCATCTCTCTCTCTCTCTCTCTCTCTCT 147545
Qy 905 aatgatctctctgctgagtggtgctccaggtatttgtggcca 944
Db 147544 TATGACTCAGTCATGTGAGCATGCAGAGGCTTCTGGCCA 147505
```

```
RESULT 13
AC022202/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-321E8 from 7, complete sequence.
AC022202
VERSION AC022202.12 GI:13992747
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 9063792
REFERENCE
AUTHORS 2 (bases 1 to 95097)
TITLE The sequence of Homo sapiens BAC clone RP11-321E8
JOURNAL Unpublished
REFERENCE
AUTHORS 3 (bases 1 to 95097)
TITLE Waterston, R.H.
JOURNAL Direct Submission
AUTHORS Submitted (26-JAN-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
REFERENCE
AUTHORS 4 (bases 1 to 95097)
TITLE Waterston, R.H.
JOURNAL Direct Submission
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (08-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 95097)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2001 this sequence version replaced gi:13431144.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0321E08

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 17 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/OTB/CHRI17, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-468B6, 200 bp overlap; the
clone sequenced to the right is RP11-561N12. Actual start of this
clone is at base position 61428 of RP11-468B6; actual end is at
base position 95097 of RP11-321E8.

FEATURES
source
1..95097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="7"
/clone="RP11-321E8"
/clone_lib="RPCI-11"
repeat_region 273..564
/rpt_family="Alu"
repeat_region 1280..1352
/rpt_family="L2"
repeat_region 1371..1404
/rpt_family="MER1_type"
repeat_region 1570..1808
/rpt_family="L2"
repeat_region 1861..2172


```
repeat_region /rpt_family="Alu"
3216..3294 /rpt_family="L1"
repeat_region /rpt_family="CR1"
3320..3610 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
4987..5293 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
5727..6021 /rpt_family="Alu"
6134..6454 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
6798..6945 /rpt_family="MIR"
repeat_region /rpt_family="MIR"
9181..9201 /rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
9239..9398 /rpt_family="L1"
repeat_region /rpt_family="L1"
9472..9785 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
9907..10485 /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
10594..10760 /rpt_family="L1"
repeat_region /rpt_family="L1"
10936..11287 /rpt_family="L1"
repeat_region /rpt_family="L1"
11288..11311 /rpt_family="(TAAA)n"
repeat_region /rpt_family="(TAAA)n"
11312..13026 /rpt_family="L1"
repeat_region /rpt_family="L1"
13027..13325 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
13326..14023 /rpt_family="L1"
repeat_region /rpt_family="L1"
14024..14328 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
14329..14394 /rpt_family="L1"
repeat_region /rpt_family="L1"
14395..14416 /rpt_family="(CAA)n"
repeat_region /rpt_family="(CAA)n"
14417..14706 /rpt_family="L1"
repeat_region /rpt_family="L1"
14736..14817 /rpt_family="L1"
repeat_region /rpt_family="L1"
14818..14837 /rpt_family="(GAAA)n"
repeat_region /rpt_family="(GAAA)n"
14838..15485 /rpt_family="L1"
repeat_region /rpt_family="L1"
15486..15770 /rpt_family="Alu"
repeat_region /rpt_family="L1"
15771..16672 /rpt_family="L1"
repeat_region /rpt_family="L1"
16682..16946 /rpt_family="L1"
repeat_region /rpt_family="L1"
17098..17162 /rpt_family="L2"
repeat_region /rpt_family="L2"
18516..18615 /rpt_family="MER1_type"
repeat_region /rpt_family="MER1_type"
18609..18650 /rpt_family="MIR"
repeat_region /rpt_family="MIR"
18773..18987 /rpt_family="L1"
repeat_region /rpt_family="L1"
19997..20162 /rpt_family="MER1_type"
repeat_region /rpt_family="MER1_type"
20282..20581 /rpt_family="Alu"
repeat_region /rpt_family="Alu"
20586..20743 /rpt_family="L1"
repeat_region /rpt_family="L1"
20832..21571 /rpt_family="L1"
repeat_region /rpt_family="L1"
21732..21798 /rpt_family="L1"
repeat_region /rpt_family="L1"
21799..22104 /rpt_family="Alu"
```

```
repeat_region 22105..22878 /rpt_family="L1"
repeat_region 22879..22902 /rpt_family="AT_rich"
repeat_region 22934..22957 /rpt_family="AT_rich"
repeat_region 23000..23270 /rpt_family="Alu"
repeat_region 23481..23523 /rpt_family="AT_rich"
repeat_region 23738..23769 /rpt_family="AT_rich"
repeat_region 24080..24144 /rpt_family="L1"
repeat_region 24145..24451 /rpt_family="Alu"
repeat_region 24452..25442 /rpt_family="L1"
repeat_region 26510..26544 /rpt_family="(TTTG)n"
repeat_region 27297..27572 /rpt_family="L1"
repeat_region 27573..27870 /rpt_family="Alu"
repeat_region 27871..28013 /rpt_family="L1"
repeat_region 28014..28324

Query Match 25.0%; Score 278.6; DB 9; Length 95097;
Best Local Similarity 58.7%; Pred. NO. 2.1e-61;
Matches 596; Conservative 0; Mismatches 404; Indels 15; Gaps 6;

QY 11 atagagatgttgaattggttattattgagaacatggcagaataattatctattctcatta 70
Db 36574 ATAAAAAATCAATAATGGTAATAATTTTGACACTGACAAAAGAACATTTGATGACACAG 36515
QY 71 gatctcttgcttttccacagatatactcttgcttttaattttcttcttaagatgatcaaa 130
Db 36514 ATGTATTTTTTTCAGGTCTTCMAAATATGTTCTTCTACATCATTTAGGATGAACACAT 36455
QY 131 ctctctgggtttttattaccatacaaaatctcttttaccacaaagccagcttttggaaattcca 190
Db 36454 TTCATATGTTATCTT----TGAAAAAGGCCCTTTTATTTTCAAGCTGGCATTTGGAATCTCG 36399
QY 191 gcaaacacacatctctctcttttccacatcttcacacttctgcttcagtcacaggtcctaag 250
Db 36398 GCAACATATTTCTTCTCTCGCACATTTTATATTTTAAAGGATCACAAAGCTTAAA 36339
QY 251 tccattgacatgataattagtcactctctctccacatactctgctctctctcactcag 310
Db 36338 AACCATGACCTGATCATCTGTCACATTGACCTTTGTTTCACATACTCATGCTAGTCATTGCA 36279
QY 311 gcaatattggtgctcttagactcttttgggttcacagaatactcaggatgatcttaggtat 370
Db 36278 GCAAAAGCTATTGTCCTCCAGATGGGTTTGAGTCACAGAATTTTCAGAATAACTTCAGATGT 36219
QY 371 aaggtcattgtctctttttaaacaaggtgatgagggcctctccatctgacacccctgectc 430
Db 36218 AAGGCTGTGTTCTACACATACAAAGCAATGAGGGCCCTTTGATCTGCACAAACCTCTCTC 36159
QY 431 ctgactgtgctccaggccat---catcagccccagcatctctctctcttggaagctcaaa 487
Db 36158 CTGAGCATGCTTTCAGGCCATCACCATTAGCCCCACACCTCTCTGGTTGGTGAGATTTAAA 36099
QY 488 catctctctgcaagtccacatcttaggattctctcttctctctctctctctcaggtctcaatgttc 547
Db 36098 CATAAAATCACAATAATACAAATATCTCGCTTTATTTATTTTGGTCCCTCAATTTTCTT 36039
QY 548 attggtgtaatactctctgctgtacactgcggctacccccagtgaaacggggccagttctct 607
Db 36038 TCACAGTGACATGATGATTATCTACATTGTAGGTTTTCAGTGTGACCCAG--ATAATTCT 35981
QY 608 gttgtcatacagcaactgtctctcttttgcctatgagctacacccacagagactgtttt 667
```



```
repeat_region /rpt_family="L1"
188..9281
/rpt_family="MERS3"
10528..10758
/Note="similar to EST A1028332 (NID:g3245641) ov90g03.xl"
repeat_region 10576..10799
/rpt_family="MaLR"
repeat_region 10813..11168
/rpt_family="MaLR"
repeat_region 11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="(TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/Note="similar to EST A1806860 (NID:g5393426) wf36b07.xl"
repeat_region 13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="TTTTTA)n"
14129..14883
/rpt_family="L1"
15223..15246
/rpt_family="AT_rich"
15290..15571
/Note="similar to EST AW445087 (NID:g6986849)"
misc_feature 15290..15571
/Note="similar to EST AW627961 (NID:g7374751) hl36d12.xl"
repeat_region 16305..16374
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="(CAAAA)n"
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="(A)n"
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="(GAAA)n"
19256..19691
/rpt_family="ERVL"
19748..19790
/rpt_family="(CAAAA)n"
20557..20613
/Note="similar to EST AW445087 (NID:g6986849)"
misc_feature 20557..20666
/Note="similar to EST A1028332 (NID:g3245641) ov90g03.xl"
misc_feature 20594..20828
/Note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
repeat_region 21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/Note="similar to EST BF899282 (NID:g12290741)"
misc_feature 22640..23050
/Note="similar to EST A1279713 (NID:g3917947) q153f03.xl"
```

```
misc_feature 22855..23086
/Note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
repeat_region 23402..23451
/rpt_family="T-rich"
repeat_region 23430..23733
/rpt_family="Alu"
repeat_region 23850..24148
/rpt_family="MER2_type"
repeat_region 24586..25150
/rpt_family="L1"
repeat_region 25151..25275

Query Match      25.0%; Score 278; DB 9; Length 65128;
Best Local Similarity 62.4%; Pred No. 2.9e-61;
Matches 552; Conservative 0; Mismatches 320; Indels 12; Gaps 7;

QY 140 ttattaccatacaaatcttcttttccacaaagcagcttttgggaatttcagcaaacacc 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12154 TGTATGTTATTAAAAAGCTGCTTGTGTTTACAAGTGTGCTGTTGGAATCTCAGTCAATACT 12213

QY 200 atccttcttcttccacatcttccaccttggtttcagtcacaggtcttaagtcattgac 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12214 TTTCGTGTTCCCTTTCAGCATCTTCACCTCTCTCTGTATATCAGACATAAAACCCACTGAC 12273

QY 260 atgataattagtcacctgtctctccacatctgctgtctctcactcagggcaaatattg 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12274 TTGATCATCTGTCCCTTGCCCTCGTCCACATAGTGAACCTCTTTCATTCGACATGTTATG 12333

QY 320 gtgtccttagacttcttgggttcacagaatactcaggatgatcttaggtataaggtcatt 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12334 CAGTCTCCAAATCTGTGTCATCATTTGAA-CTTCAGAAATGACTTCAAGTGAAGGACATT 12392

QY 380 gtcttttaaaagaagtgatgaggggctctccatctgcaacctctgctctgagtg 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12393 TTCCACCTGAGCAGGGGTGTGTGGGGCTCTCTATCTGCACCACCTGCCCTCGAGTGTG 12452

QY 440 ctccagg---ccatcatcagcccgagcatctctctgtggcaaaagctcaaacatctctt 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12453 TTCCAGGCCACCAACCATCAGCCCCAGCACCCTCTGATTTGGCAAGGTTTAAACATAAATCT 12512

QY 497 gcaatcacatcttaggtattcttcttcttcattgggttcacacatgttcatgtt-gt 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12513 TCNAATTACATTTATCATGTTTCTTCTTCTTATGTTCTCTCAACTTGTCTTTTCAGTAGT 12572

QY 556 aatctctgtgtacactcggtctacccccagtg-aaacggggccaggtcttctgttgc 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12573 AACAGGATCTTCTACACTGTAGTGTCTTCCAGTGTGAGCAGACAAAATGACTGAATATA 12632

QY 615 atacagcactgtctcttcttcccatgagctacacccacagagactgtttttcacacta 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12633 GGTAAGTACTGCTCAGTGTTCCTCCATGAGCTTCATCTCAGGGCCACGGTTCTTATTCTG 12692

QY 675 atgacttgagggtatgacaccttttatagggttcctctcagagggtacatgggtg 734
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12693 ACAACACTTAGGGATGCTTTCCTGGTAGGAATACGTCTTCTTAATTGCATACATAGTGTG 12752

QY 735 attattttatcacagaataaagagctctctcagtccttcacgagccagcagctgtcccc 794
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12753 ATCTCTTATTTCAGGC-CTCAGAGGTGATCCCAACCACTTCACAGCAATGGTTCCTTCCC 12811

QY 795 --gagcttcaccagtgaagaagcctcccaggtcttcttactgtggtgagttt---tgt 849
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12812 AGGACTTCCCCTGAGAAACGGGCCATCCAGACCATCTGCTGCTGGTAAGTTCCTCTCT 12871

QY 850 cttcacactgggtggactttacgttcttcttcttccatttcagggaggtgtgacatgataaata 909
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12872 GGTCTCTCTGTTGGTGGATTTCATCATCTCTCTCTCTCCTCAGCAATGTTGTGGGCATATGG 12931

QY 910 tctctgtagtggtggtccaggttattgtggcccaataagctatgcgcgaattagctcttt 969
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12932 CCCAATCAACCA3CAAGTTTCAGAGTCTTGTAGTCAATGCTATGCGGTGATCAGTCTCT 12991

QY 970 gatgctaatttatgctgataaccaaaatattcaagactctgcaaa 1013
```


Db	15886	ATTAAATGATGAAGTTGGTAATAATTTTGACTACACAATAACTTAACATTTGATGATGAAGA	15945
Qy	71	gatctcttgcttttccacagatgac-cccttgctttaaatttctctctaagatgatcaa	129
Db	15946	GGTCTATTGCTTTTTCAGGTATATAAAATATAAATCCCTCTTCATCATTAAGACGACCA	16005
Qy	130	acttccctgggtttattaccatacaaaactctctttttatccacaagccagctttggaaatttc	189
Db	16006	TGTTTCCAGTTAAGGTACTACTAACAATTTGGCCTTTCTTTTCAAGCTGGTATAGACACTAG	16065
Qy	190	agcaaacaccatccctctcttctttccacatcttcacaccttggtttttcagtcacaggtctaa	249
Db	16066	AGCCAACAACCTCTCTCTATTCTTCCAATACTTCTCACTCCTTCAGGATCACAGGCCATA	16125
Qy	250	gtccattgacatgataattagtgccacctgtctcttcacatctcctcactgtctcttcactca	309
Db	16126	CCCCACATGACCCCATCACTGTCTACCTGGCCCTTTGTCCACTTAGGGATGTCCTCATGTGT	16185
Qy	310	ggcaaatagggtcccttagactcttggtttcacagaatacactcagatgatcttagtata	369
Db	16186	GGTCTCTTGGCATCTCCAGACCTGGTTTGAGTCACTGTATTTTCAGAATGACTTCAAGTG	16245
Qy	370	taaggctcattgtctttttaaacaagggtgtagaggggctctccactctgcacccctgcct	429
Db	16246	TAAGGCATTCTTCGTGCATGCACAGGGTGATCAGAGGCTCTCCATCTGCACCACCTGCCT	16305
Qy	430	ccttgagtgtcctcaggc---catcatacagccccagcatcttctccttggcaagctcaa	486
Db	16306	CCTGAGCATGCTCCAGGCTGTGGCCATCAGCCCTGGCACCTCTCTGGTGTGCAAGGATTAA	16365
Qy	487	acatccctctgcaagtcacatcttttaggattcttctcttctcattggtctcctcaacatgt-	545
Db	16366	ACAGAAATTCNAAGGTTTACATCTTTCACTCTCTCTTCTATGTTGGGTCTTCAGTTTGTG	16425
Qy	546	tcatttggtagtaactctctgtctgacactgoggctacccccagtgaaacg9ggccagctctt	605
Db	16426	TCTCAGTAGTAACCTGCTCTCTCCACTGTGGCTTCTTAATGGGACCAAGACGGTTGT	16485
Qy	606	-ctgttttgtcatacagcacgtctccttttggccatgagctacacccacagagagactgttt	664
Db	16486	GCTAAGTATCAGTAAATACTGATCACTTTCTTCCATAAGCTACATTAATCAGGAGCCCTGTC	16545
Qy	665	tttcacactaatgactttgagggtatgcaccttatagggttctatggtctctctcaagagg	724
Db	16546	TTTCATGCTGCCACTGCTTAGGAATGTCTTCTTTGACGAATCATGAGCCCTCAAGTGC	16605
Qy	725	ctacatggtgatattttatcacagacaataagaggctatctcagtgcccttaacagcaca	784
Db	16606	ATACATGGTAAATTCTCTGTGTTCAGGC-ATCAGAGACAATCCCAATACCTTCATAGAACCA	16664
Qy	785	gccgtgccccgagctcaccaagtgaagaagagccctccagagctatcttactgctggtgaagt	844
Db	16665	ACCTCTCTCCAAGAGCCTCCCCAGAGAAAGGGGACCA--TCTTTGTTGGTGAGTTC	16722
Qy	845	tttgtcttcacatactgggtggactttacgtctctcatctttagcagaggtgtgacatggata	904
Db	16723	TTTCTGGTGCATGTACTGGGGGACCTTATTATCTCATCTCTCTCAACTCTATTATTGGCT	16782
Qy	905	aatgattctctcgtagtgtggtctccaggttatttggccca	944
Db	16783	TATGACCTCAGTCATTTGTGAGCATGCAGAGGCTTCTGGGCA	16822

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 02:40:43 ; Search time 2520.32 Seconds
(without alignments)
5965.754 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatgagatgt.....gagaggttaatccatcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estma:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	199.6	17.9	725	12	AQ078448
2	186.6	16.8	690	12	AZ840584
3	186.2	16.7	777	12	AZ741291
4	159.6	14.3	808	12	BH039878
5	153.2	13.8	729	12	BH083291
6	142.6	12.8	545	12	AZ748309
7	141.8	12.7	617	12	AZ795382
8	136.4	12.2	635	12	BH062438
9	133.8	12.0	574	9	A1806860
10	130.4	11.7	610	12	AZ222872
11	130	11.7	467	12	AQ117475
12	129.4	11.6	615	12	AZ553624
13	129	11.6	563	12	AZ810472
14	128.4	11.5	740	12	BH267496
15	127	11.4	638	12	AZ553591
16	120.6	10.8	853	12	AZ265296
17	119.8	10.8	493	12	AZ759046

18	116.4	10.4	447	9	AA012849
c 19	116.4	10.4	768	12	BH060668
c 20	115.4	10.4	597	12	AZ422002
c 21	115	10.3	715	12	AG013852
c 22	113.2	10.2	698	12	BH039875
c 23	112	10.1	643	12	BH048022
c 24	111.8	10.0	397	9	AA021085
c 25	111.8	10.0	586	12	AZ093307
c 26	111.2	10.0	424	10	H86939
c 27	110.2	9.9	717	12	AG013832
c 28	109.6	9.8	682	12	AG130179
c 29	105	9.4	511	12	B67083
c 30	104	9.3	495	12	AQ223006
c 31	100	9.0	243	12	AZ737781
c 32	98.4	8.8	716	12	AG013834
c 33	98	8.8	696	10	BE968588
c 34	97.8	8.8	715	12	AZ994117
c 35	97.2	8.7	725	12	AG013831
c 36	94.6	8.5	382	9	AA442630
c 37	94.4	8.5	677	12	BH267500
c 38	92.6	8.3	580	9	AW973537
c 39	90.4	8.1	481	12	B92737
c 40	88	7.9	593	12	AZ754210
c 41	86.8	7.8	597	12	AZ896363
c 42	86.2	7.7	406	12	AZ767149
c 43	85.6	7.7	658	12	AQ666848
c 44	84.6	7.6	207	12	B59457
c 45	83	7.5	675	12	AG000745

ALIGNMENTS

RESULT 1

AQ078448

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AQ078448 725 bp DNA linear GSS 20-AUG-1998
CIT-HSP-2358C21.TFB CIT-HSP Homo sapiens genomic clone 2358C21, DNA
sequence.
AQ078448 GI:3439632
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 725)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other_GSSs: CIT-HSP-2358C21.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 338 0200
Fax: 301 338 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2358C21"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"

Db 390 ATATCAAGTGCATTTCTTCTGAGTCTCCTCTATATGTTAATTGGCAGTCAACTCTTG 449

Qy 565 ctgtacactcggctaccccccgagtgaaacggggcgagctctctctgtt-tgtcatacagac 623

Db 450 GTATCCATCATTTGGCACCACCCCAATTTGACCATGAATGATTTATTTATGTTACTACGCC 509

Qy 624 tttctcttttggccatgagctacaccacagagagactgtttttcacacaaatgactttg 683

Db 510 TGCTCTATTCTACCTTGAGTACTCATGCAAGACATATATTCTACACTGCTAGCCATC 569

Qy 684 agggatgcacctttataggggttcctctcaagagagctacacatggtgattatttta 743

Db 570 AAGGAATTTCTTCTATTAGTCTCATGCTCTCTAATTTGGTACATGGTAGCCCTCTTA 629

Qy 744 tacagacaataagaggctatctcagtccttcacacagcagcagctgccccga 796

Db 630 AGCATGC-ACAGGAACAGACAGCATCTTCAATGGAACCAACCTGTCCCCCA 681

RESULT 3

AZ741291/c 777 bp DNA linear GSS 25-JAN-2001

LOCUS AZ741291

DEFINITION RPCI-24-90L5.TV RPCI-24 Mus musculus genomic clone RPCI-24-90L5, DNA sequence.

ACCESSION AZ741291

VERSION A2741291.1 GI:12516500

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 777)

AUTHORS Zhao S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-24-90L5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 90 row: L column: 5
Seq primer: SP6
Class: BAC ends.

FEATURES

source

1. .777
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-90L5"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 260 a 141 c 154 g 222 t

ORIGIN

Query Match 16.7%; Score 186.2; DB 12; Length 777;
Best Local Similarity 61.1%; Pred. No. 2.1e-34;

Matches 389; Conservative 0; Mismatches 238; Indels 10; Gaps 5;

Qy 385 tttaacaaagtgatgaggggctctccatctgcacccccgcctcgtgagtgctccca 444

Db 774 TTTAAACAGGGCAATGAGAGGTCCTCTATACAGCATCACCTGCCTCTAGAGTGTGTC 715

Qy 445 ggcctatc--atcagccccagcatcttctcttgcaagctcaaacatctcttctgcaag 501

Db 714 GGCTGTGACGATCAGTCCCAACTCCGCTTTGTTGGCAATTTAAACATAAACTAAAA 655

Qy 502 tcacattttaggattctctcttctctcatgggtcctcaacatgttcttctggtgtaattct 561

Db 654 ATACATGATCTATTCTTCTTCTATTGCTCTTCAATTTGTCATTCAGCATTAACAG 595

Qy 562 ctgct--gtacactgcggtacccccagtgaaacggggcgagctctcttctgttgcataca 619

Db 594 GATCTTCTATACCTGGTCTTATATCAACGTCGTCGTCGACACCAACGAGGTGAAGGTC 535

Qy 620 gcactgtctcttttgcctatgagctacacccacagagagactgtttttcacactaatgac 679

Db 534 ATACTGCTCACTCTTCCCTATGAACAACATCATCAGGGGACTAATTTCAACAATCAAC 475

Qy 680 ttgagggatgtcacctttataggggttcctcctcctcaagagggctacatcgtgattat 739

Db 474 CTTAAGAGATCTATTTTGTAGGAGTCATGCTGAGCACAAGTACATACATGCTGATTAT 415

Qy 740 ttatcacagacaataagaggtctctcagtccttcacacagcagcctgtcccccagtc 799

Db 414 CTTCTTTCAGAG-ATCAGAGGCAATGCAAGCATCTTCATAGCGACAGCCACTTGA-GAG 357

Qy 800 tcaccagtgaagagagctcccgagctctctactctctggtgagttt---tgtcttcaca 856

Db 356 TCCCTCAGAAAAAGAGCCACCCACCATCTTGTCTGCTGTTTCTATGCTGGTCATG 297

Qy 857 tactgggtgactttacgttctctcattttctcagggaggttgacatggataaagtctctgt 916

Db 296 TACTGGGTGGACTTTATATGTCATCTAGAACAGTCTTTTATGATGATGATGACCCAGTC 237

Qy 917 ctagtgggtccaggtattgtggccaaatagctatccgcaatagctccttcttctgacta 976

Db 236 ATGCTTACTGTTTCAGAAAGTTGTGTTGAATGCCCTATCCCAATTACTCCTTTTAACAA 177

Qy 977 attatgtctgataacaaatattcaagactctgcaa 1013

Db 176 ATTAGTTCTGATATCGAATACTCATGATTTCTAAAAA 140

RESULT 4

BH039878

LOCUS BH039878

DEFINITION RPCI-24-273F22.TV RPCI-24 Mus musculus genomic clone RPCI-24-273F22, DNA sequence.

ACCESSION BH039878

VERSION BH039878.1 GI:14818265

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 808)

AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-24-273F22.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html Plate: 273 row: F column: 22
Seq primer: T7
Class: BAC ends.

```
FEATURES
source      Location/Qualifiers
1..808      /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone_lib="RPCI-24-273F22"
            /clone_lib="RPCI-24"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
            RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pTARBAC1 cloning vector at the
            BamH1 sites using MboI partially digested male C57BL/6J
            DNA."
BASE COUNT  234 a 172 c 139 g 263 t
ORIGIN
Query Match      14.3%; Score 159.6; DB 12; Length 808;
Best Local Similarity 55.9%; Pred. No. 5.6e-28;
Matches 389; Conservative 0; Mismatches 294; Indels 13; Gaps 4;

QY 328 agactcttttggttcacagaataactcaggatgatctcttaggtataaagtgctctttt 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39  AAACATATTAGAGTCACATAAAATTTGGGAATGGCATCAAAATGTAACAACTTTTCCAT 98

QY 388 aaacaaggtaagggcctccactctgcacccctgctcctcctgagtggtcaccaggc 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99  AAACAGAGGATGAGAGGCCCTCTCTATCTGCATCACCTCGCTTGTGAGTGATATCCAGGC 158

QY 448 catc---atcagcccgagcatctctctcttgccaaagctcaaacatcctctcgcaagtc 504
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TATCCCAATCAGTCCCGTAGCTCTTTGTTGGCAAAATTTAAGAGTAAAGTAAACAAACA 218

QY 505 catcttaggattctctctctctctctctctcctcctcctcctcctctctctctctg 564
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 GATCATCTAGCGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 276

QY 565 ctgtacactcggctaccccccagtgaaacggggccagctctctctgtgtctacacagcact 624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 ATCTTCCATGTTGCTGATTTTCAATGTGAAACACTACAGATGAAGGTCACTAAATCCT 336

QY 625 gttccttttgcacatgagctacacacacagagagactgttttttcacactaatgactttga 684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GCTCACTCTTCCCGCATGAACATACATCATCAGGGGGTGTGATTTAAACAGTGCACATGTCCA 396

QY 685 gggatgtcacctttatagggttcattgggtctctcctcagaggctacatgggtgattatttat 744
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 GAGATATATCTGTGTAGGAGTATGCTGTATCACAAGTACATACATGCTGGTGTATCTTAT 456

QY 745 acagacaataaagggtctatctcagtcgcttcacacagccagcgtctcccgagctcacc 804
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 GCAACATCAGCAATGCAACGATTTTCATAGCATCAGCCCA-----CCTGAGACATCTCC 511

QY 805 agtgaagaagagctcccgagctctactactgctgggtgagt---tttgcttcacactg 861
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 512 TGAGAAAAGGGCCACCAGACCATCTTGCTGTAGTAGTTTCTTCTGTGTGTGTACTG 571

QY 862 ggtggaactttacgtctctcatttccaggagggtgtgacatggataaattcctctcgtagt 921
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 572 GTTGGGACTTCATCATCTCAATCACTCAATCAATCAATCAATCAATCAATCAATCAATCTCT 631

QY 922 gtggctccaggttatgtggccaatgactatgccgaattagtccttccatgctaatatta 981
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 632 GACCATTCAGACGTTGTGTATGTATGCCTATCCCAAAATTTGCTCTTTGTGTACAGCTCAG 691
```

```
QY 982 tgctgataacaaatattcaagactctgcaaatgtt 1017
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 TTCTGATAGAGAGAAATATCGTATACTATAAAAAATTT 727

RESULT 5
BH083291      Location/Qualifiers
LOCUS      RPCI-24-309J17.TV RPCI-24 Mus musculus genomic clone RPCI-24-309J17
DEFINITION      , DNA sequence.
ACCESSION      BH083291
VERSION      BH083291
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 729)
AUTHORS      Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aklnret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-24
JOURNAL      Unpublished (1999)
COMMENT      Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the mouse BAC library RPCI-24. For BAC
              library availability, please contact Pieter de Jong
              (pdejong@mail.cho.org). Clones may be purchased from BACPAC
              Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
              page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
              Plate: 309 row: J column: 17
              Seq primer: T7
              Class: BAC ends.
FEATURES
source      Location/Qualifiers
1..729      /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone_lib="RPCI-24-309J17"
            /clone_lib="RPCI-24"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
            RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pTARBAC1 cloning vector at the
            BamH1 sites using MboI partially digested male C57BL/6J
            DNA."
BASE COUNT  170 a 181 c 138 g 240 t
ORIGIN
Query Match      13.8%; Score 153.2; DB 12; Length 729;
Best Local Similarity 58.2%; Pred. No. 1.9e-26;
Matches 384; Conservative 0; Mismatches 263; Indels 13; Gaps 6;

QY 78  tgcctttctccacagatctcttgccttaattctctcctcctcctcctcctcctcctg 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74  TGTCTTCTTCCTCTGTCACCTGTTCTCACATACATGAATAAGACAGCAGACATACATA 133

QY 138 gttttattaccatacaaaattctctttattatccacagccagcttctggaatttcgcaaaaca 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 CTAAATTCTAACATTTGGGAAC-ACATTTTCTCTGATTTTCTGTTGGCATCACAGAAACA 192

QY 198 ccatcctctctctttccacatctccacttctgtttttctcagtcacaggtctcaagtcattg 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 GCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 252
```

QY 258 acatgataattagtcacccgtgtctctcatccacataactgctgtctcttctcactcaggcaatat 317
Db 253 ACTGCCCATGCTCTCTCTCCCTAAATCCACCTCTGATGCTTCT---GGTCGAGCAT 309
QY 318 tgggtccttagactcttctgtgtccacagaataactcagagatgctcttaggtataaggta 377
Db 310 TCATAGCCACACACATTTTATCTCTCGGAGGGCTGGGATGACATCATATGTAAATTC 369
QY 378 ttgtctttttaaacaggtgtaggggctctcctcatctgcaccccccgcctcctcagtg 437
Db 370 TTGTGTACCTGTACAGAGTTTGTAGGGGTCTCTCGCTTTGACACACCATGTTGAGTG 429
QY 438 tctcagagcatca---tcacccagcagcatctctcttgcgcaagctcaaacatcctt 494
Db 430 TCTTCAGGCCCATCATCTTATCTTCCAGAGCTCTCTGTTTATCAAGTTCAAGCATATAT 489
QY 495 ctgcaagtcacatcttaggattctctcttctccttggtggtcctcaacatgttctt 553
Db 490 CTCTTCATCATCTTATGTGCTATTTCTTCTGAGTGTCTCTATATGTTAAATAGCA 549
QY 554 gtaactctctgtgtacactgagggctacccccccagtgaaac---ggggccagctcttctgt 609
Db 550 GTCAACTCTTGTGTATCAATTCATTTGCCACCCCTAATTTGACCAAAATGACCTTCTCTACTT 609
QY 610 ttgtcatcagcagctgttctcttcttcttctccttggtggtcctcaacatgttctt 668
Db 610 ATGTTACTCAGTCCCTGCTATCTATCTTCTGAGTGTCTCTATATGTTAAATTC 569
QY 669 acactaatgactttgagggatgtcacctttatagggttcattggtcctcagagagctac 728
Db 670 AGCGTGTACGCATCAGGGGATCTTCTTATTTAGTCTCATGTTCTCTCTCAACTGGGTAC 729

RESULT 6
A2748309
LOCUS
DEFINITION A2748309 545 bp DNA linear GSS 25-JAN-2001
A2748309
ACCESSION A2748309
VERSION A2748309.1 GI:12533043
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-63G18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 63 row: G column: 18
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-63G18"

FEATURES
source
Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-63G18"

/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 160 a 116 c 112 g 157 t
ORIGIN
Query Match 12.8%; Score 142.6; DB 12; Length 545;
Best Local Similarity 61.1%; Pred. No. 6.1e-24;
Matches 334; Conservative 0; Mismatches 204; Indels 9; Gaps 6;
QY 301 ctctcactcaggcaataattggtgtccttagactcttcttgcacagaataactcagatga 360
Db 1 CCTCACTGGAGGAAATGTCTATGCTTGCAGACATATTTGAGTCACCTGAATGTTGAGAATGA 60
QY 361 tcttaggtataagggtcattgtctcttttaacaagggtgtaggggctctcactatcgcac 420
Db 61 CATCAAAATGTAAGGCAACTTTATACAAAACAGGGGTGATGAGAGGCTCTCTATACGACAT 120
QY 421 cccctgcctcctgagtgctcaggc---catcatcagccccagcatctctccttggc 477
Db 121 CACCTGCCCTCCCTGAGTGTGATCCAGGCTGTCAATATCATCTCCGAGTACCTTTATGGTGCA 180
QY 478 aaagctcaa-acatcctctctgcaagtccatctcttaggattcttcttctctcatgggtcc 536
Db 181 AAATTTAAACAATAAATAAGAAAACACATGGTCAATGCTTCTTTTATATATTTGGTCT 240
QY 537 tcaacatgt--tcattggtgtaattctctgtgtacactgcgggtacccccagtg--aaac 593
Db 241 TTCAATATTTTCTTCTAGTAGTAGTCTCTGATCTTCTATATACCTGGTGTATACCAATGTGAGT 300
QY 594 gggggcagctctctgtttgtctacagcactgtctcttcttgcctatgagctacaccccac 653
Db 301 GAGACCAACAGCATGAGATGACTAAGTCTCTCAATCTTGCCTCAATGAACTACATCATC 360
QY 654 agggactgtttttcacacataatgactttgagggtatgtcacctttatagggttcattc 713
Db 361 AGGGAATGGTTGTAACAGTACACACCGTCAGAGATGTGTTCTTCTGGGAGTTATGCTG 420
QY 714 ctctcaagggtacataagtgattatttttatcacagaataagaggctatcctcagtgct 773
Db 421 ATCAACAAGTGCAATACATGTTGATATCTTTGTCAGAC-ATCAGAGGCAATGCAAGCATCT 479
QY 774 tcacacagcagcctgtcccgagtcctcaccagtgaaaagagcctccaggtctatctac 833
Db 480 TCATAGCATCA-ACAACCTGAGAACCTCCCTTGAGAAAAGGCCACACAGACCATCTTGC 538
QY 834 tgcgtgt 840
Db 539 TGCTAGT 545
RESULT 7
A2795382/c
LOCUS
DEFINITION 2M0049N06R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGC2M0049N06 R, DNA sequence.
ACCESSION A2795382
VERSION A2795382.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 617)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahnoun,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

Best Local Similarity 63.2%; Pred. No. 2e-22;		Matches 259; Conservative 0; Mismatches 146; Indels 5; Gaps 3;	
QY	607	tgttgtcatacagcactgttccttttggccatgagctacaccacagagagactgtttt	666
Db	3	TGAAGGTGATTAATACTGCTCAGCTCTTCCCATGAAGTACATCATCAGGGAGCTGATTT	62
QY	667	tcacactaatgactcttgagggatgctacacctttatagggttcatgtgctctccaagagct	726
Db	63	TAACCATGACAACCTTAAGAGATCTGTTCTTGTAGGAGTCTGCTGACTACGAGCAGAT	122
QY	727	acatgggtattttttatacagacacaaagaaggctatctcagtcagtccttcacacagcagc	786
Db	123	ACATGGGTGATATCTTGTTCAGAC-ATCAGAGGCAATGCTTTTCTTCATAGCATCAAC	181
QY	787	ctgccccagagctccacagtgaaagagcctccaggctatcttactgctgggtgagt--	844
Db	182	C-ACCTGAGAGCCTCCCTGAGAAAGGCCACATCATCAGTCTTGTCTGGTGGTGT	240
QY	845	-ttgttcttcacatactcgggtgagctttaaogttctctatcttcaggaggtgagcatggat	903
Db	241	CTTTGTGGTCTACTGCTGCTGGTGGATCTTCATCATCTCATCCACCTCAGTCTGTATGGAT	300
QY	904	aaatgattctctgctatgtggtctccagggttatgtggccaatagctatgcgcgaattag	963
Db	301	GTATGACCCAGTCACTCTGACTGTTTCAGAGATTTGTGATGAATGCTTACTTCAATATAC	360
QY	964	tcctttgatgctaattatgctgataacccaatttcaagactctgcaaa	1013
Db	361	TCCATTGGTACAAATCAGTCTCTGATACCGAATAATCATGATGCTGAATAA	410
RESULT	9		
LOCUS	A1806860/c		
DEFINITION	wf36b07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357653 3' similar to TR:Q62B55 Q62B55 PHEROMONE RECEPTOR VN6. ;, mRNA sequence.		
ACCESSION	A1806860		
VERSION	A1806860.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 574)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb@email.nih.gov This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 659 Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 447. Location/Qualifiers 1..574 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2357653" /lab_host="DH10B" /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of		
FEATURES	SOURCE		

RESULT 12	AZ553624	615 bp	DNA	linear	GSS 20-NOV-2000
LOCUS	RPCI-23-209L1.TV	RPCI-23	Mus musculus	genomic clone	RPCI-23-209L1, DNA sequence.
DEFINITION	AZ553624				
ACCESSION	AZ553624.1	GI:11233124			
VERSION					
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	1 (bases 1 to 615)				
	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroi, M., de Jong, P. and Fraser, C.M.				
TITLE	Mouse BAC End Sequences from Library RPCI-23				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: RPCI-23-209L1.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html Plate: 209 row: L column: 1 Seq primer: T7 Class: BAC ends.				
FEATURES	source				
	1..615				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCI-23-209L1"				
	/clone_lib="RPCI-23"				
	/sex="Female"				
	/lab_host="DH10B"				
	/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."				
BASE COUNT	178 a 128 c 107 g 202 t				
ORIGIN					
	Query Match 11.6%; Score 129.4; DB 12; Length 615;				
	Best Local Similarity 58.3%; Pred. No. 9.8e-21;				
	Matches 284; Conservative 0; Mismatches 196; Indels 7; Gaps 3;				
QY	99	tttgcttaatttcttcttaagatacaaaacttctctggtttttattaccatacaaatct	158		
Db	80	TATGACTCAACAATGCTCTTAAGATAAATTCACCTCTCACTTATGTCTCATTTGGAGAAT	139		
QY	159	tct-tttatccacgaagccagcttggaatttcagcaaacaccatctcttctttccac	217		
Db	140	GCTCTTATATCCAACTGGGTAGAGAGTCCTAGCTAATATGTGCTTCTTGTTTCTAT	199		
QY	218	attctcacctttgttttcagtcacaggttcagtcacatgacataaattagcacctg	277		
Db	200	ATTTTCA---TAATCTTAGGGCAGACACACCTAAGCCCATGACCTAATCTCTCTCAACAG	256		
QY	278	ttctcatcacactgctgctcttctcactcaggcaaatattggtccttagacttcttt	337		
Db	257	ACTTTTCATTCACATAATGCTGTTCTTCTTCACGAGGAGATATTTTGATACAGATATTTT	316		
QY	338	ggttcacagaactcagatgatcttagtataaagtcattgtctttttaaacagggtg	397		
Db	317	GAGTCAATGAATATTTGAGAATGACTTCAAATGCAAAACAACATTTTACATATGACGGTA	376		
QY	398	atgaggggcctctccatctgcacccctgctcctcctcagtgctgctccagge---catactc	454		
Db	377	ATGAGAGGGCTCTCTATCTGCACCACCTGCCTCTGAGTGTGTCCAGGTGTCACCATC	436		
QY	455	agccccagcatcttctctctggcgaagctcaaacctctctcgaagtcacatcttagga	514		
Db	437	AGTCCCAATACCTTCGCTGTTGGCAAAATTTAAACATAAAATAAAATACACTATCAA	496		
QY	515	ttctctctttctcagtggtctcacaatgctcattggtgtaattctctctgctgacactg	574		
Db	497	TGCTTTCTCTATATTTGGGTCTTTAAATTTGCTTCTCAGTAGTAACCTCATCTCTATG	556		
QY	575	cggctac	581		
Db	557	TTGGTGC	563		
RESULT 13	AZ810472	563 bp	DNA	linear	GSS 20-FEB-2001
LOCUS	2M0076E06F	Mouse 10kb plasmid	UUGC1M library	Mus musculus	genomic clone UUGC2M0076E06 F, DNA sequence.
DEFINITION	AZ810472				
ACCESSION	AZ810472				
VERSION	AZ810472.1	GI:12977756			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 563)				
	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D. Weiss, R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0076 row: E column: 06 Seq primer: CGTGTAAACGACGCGCCAGT Class: plasmid ends High quality sequence stop: 563.				
FEATURES	Location/Qualifiers				
	1..563				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="UUGC2M0076E06"				
	/clone_lib="Mouse 10kb plasmid				
	/sex="Male"				
	/lab_host="E. Coli strain XL10-Gold, Ti-resistant, P-"				
	/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA				

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

BASE COUNT      164 a      116 c      105 g      178 t
ORIGIN

Query Match      11.6%; Score 129; DB 12; Length 563;
Best Local Similarity 61.6%; Pred. No. 1.2e-20;
Matches 257; Conservative 0; Mismatches 155; Indels 5; Gaps 3;

QY 613 tcatacagcactgttccttttggccatgagctacacccacagagagactgtttttcacac 672
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 9 TCACATAATCCGCTTACTCTCCGCACTCACTCCATCATCAGGGTATTTCATTTTACAG 68

QY 673 taatgactttgagggtatcaacttttatagggttcattggtcctctcacaagaggtacatgg 732
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 69 TGACAACCTCCAGAGATGATTCTTGTAGGAATGATGCTTAACCAACAGATACATACATGG 128

QY 733 tgattattttatcacagaataagaggtatctcagtcgcttcacacacagccagcctgtcc 792
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 129 TGATATCTTATTATAGAC-ATCAGAGGCGAGTGCACAGCACCTTCATAGCATCAAGC-ACCT 186

QY 793 ccgagctcaccagtgaagaagcctccagcgtatcttactcgtcgtgagtc---tttgt 849
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 187 GAGAGCATCCCTGAGAAAGAGCCACAGACTATCTTGCTGCTAGTGATTTCTTTGCT 246

QY 850 cttcacactgggtgagctttacgtttctcattttcagggaggtgtgacatgataaatga 909
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 247 GGTCTGTACTGGTGGAGCTTCATCATCTCAACCATATCAGTCTCTGTTGGAGGTATGA 306

QY 910 ttctctgtagtggctccaggtattgtgccaatagctatgcgcgaattagtccttt 969
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 307 CACAGTCATCCTGAGGGTTTCAGAAAGTTTTTGTATGATATGCTATGCCACAAATAGTCCCTTT 366

QY 970 gatgtaatttatgctgataacaaatattcagaactctgcgaattgttggtttaa 1026
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 367 GGTACAAATCAGTCTCTGATAACAGATATAATATTATGGTGAAAACATGTACTCAA 423

```

```

RESULT 14
LOCUS      BH267496/c
DEFINITION CH230-82L13.TVB CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH267496
VERSION    BH267496.1
KEYWORDS   GSS.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 (bases 1 to 740)
AUTHORS    Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
            A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
            Jong, P. and Fraser, C.M.
TITLE      Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL    Unpublished (1999)
COMMENT    Other_GSSs: CH230-82L13.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research

```

```

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
plate: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
Size: 82 row: L column: 13
Seq primer: 77
Class: BAC ends.
FEATURES             Location/Qualifiers
     1..740
     /organism="Rattus norvegicus"
     /strain="BN/SSNHsd/MCW"
     /db_xref="taxon:10116"
     /clone="CH230-82L13"
     /clone_lib="CHORI-230 Segment 1"
     /sex="Female"
     /cell_type="Brain"
     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong."
BASE COUNT      250 a      136 c      136 g      218 t
ORIGIN

```

```

Query Match      11.5%; Score 128.4; DB 12; Length 740;
Best Local Similarity 64.9%; Pred. No. 1.8e-20;
Matches 253; Conservative 0; Mismatches 131; Indels 6; Gaps 4;

QY 627 tctttttgccatgactacacccacagagagactgtttttcacactaatgactttgag 686
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 740 TCACCTCTTTCCCATGAACATCATCATCAGGGGACTGATTTTACAAATGACAACTCCAGA 681

QY 687 gatgtcacctttatagggttcattggtcctctcaagagggtacatggtgattattttat 746
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 680 GATGTTTCTTA-GTAGAGTGCATGCTGCACAAACAGCATGTACATGGTGATTATCTTGTGC 622

QY 747 agacaataagaggtatctcagtcgcttcacacagcagcctgtcccccaggtccaccag 806
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 621 AGAC-ATCAGAGGCGAGTGCAGCATCTTTCATAGCATCATCACT-TGAGAGCATCCCTG 564

QY 807 tgaaaagagcctccacaggtctattctactcgtcgtgag---tttgtcttcacatactgg 863
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 563 AGAAAAGGGCCACCCAGACCATCTTCTGTTGGTGGGTTTTTTTGGTGCATGACTGGG 504

QY 864 tggactttacgttctcatttttcaggaggtgtgacatggataaatgattctctgctagtgt 923
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 503 TGGACTTCATCATCTCATCCACCTCAGTCCTATTATGGATGTATGACCCAGCATCCTGA 444

QY 924 ggtccaggttattgtgccaatagctatgcgcgaattagctcctttgagctaatattatg 983
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 443 CTGTTTCAGAAAGTTTGTGATGAATGCCCTATCCCAAAATTACTCCTTTGATACAAATCAGTT 384

QY 984 ctgataaccaaataattcagaactctgcaaa 1013
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 383 CTGATAACCGAATAATCATTCGCTGAAAA 354

```

```

RESULT 15
LOCUS      AZ553591
DEFINITION RPCI-23-209J1.TV RPCI-23 Mus musculus genomic clone RPCI-23-209J1,
            DNA sequence.
ACCESSION AZ553591
VERSION    AZ553591.1
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


REFERENCE 1 (bases 1 to 638)
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: RPCI-23-209J1.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html

Plate: 209 row: J column: 1

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .638
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-209J1"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

184 a 135 c 105 g 214 t

BASE COUNT
ORIGIN

Query Match 11.4%; Score 127; DB 12; Length 638;
Best Local Similarity 59.3%; Pred. No. 3.8e-20;
Matches 272; Conservative 0; Mismatches 180; Indels 7; Gaps 3;

QY 99 ttgtcttaatttcctctcaagatgatcaaaacttcctggtttttattaccatacaaatct 158

Db 39 TATGACTCAACAATGGCTTTAAGATAAATTCACCTCTCACCTTATGTTCTCATTTGGAGAAAT 98

QY 159 tct-tttatccacaagcagcttttggaaatttcagcaaacaccatccttctcttttccac 217

Db 99 GCTCTTTATATCAAGCTGGTTAGGAGTCTAGCTAATATGTCGTCTCTGTTTCTAT 158

QY 218 atttccactttgttttcagtcaggtctaaagtcacattgacatgataatagtcacctg 277

Db 159 ATTTTCA---TAATCCTAGGGCAGACACCTAAGCCCATGGACCTAATCTCCTGTCAACAG 215

QY 278 tctctatcaacaatactgctgtcttcactcagggcaaatattgggtgcctcttagactctttt 337

Db 216 ACTTTTCATCACATAATGCTGTGTTCTTTCACCTGCAGAGATATTTTGCATACAGATATTTT 275

QY 338 ggttcacagaatactcaggtatgctcttaggtataaggtcattgtcttttttcaacaagggtg 397

Db 276 GAGTCAATGAATATTGAGAATGACTTCAATGCAAAACAACACTTTTACATATGCAGGGA 335

QY 398 atgaggggctctccatctgcacccctcctcctctgagtggtgctccaggc---catcatc 454

Db 336 ATGAGAGGCGCTCTCTATCTGCACCACTGCCTCCTCAGTGTGTTCCAGGCGTGCACCATC 395

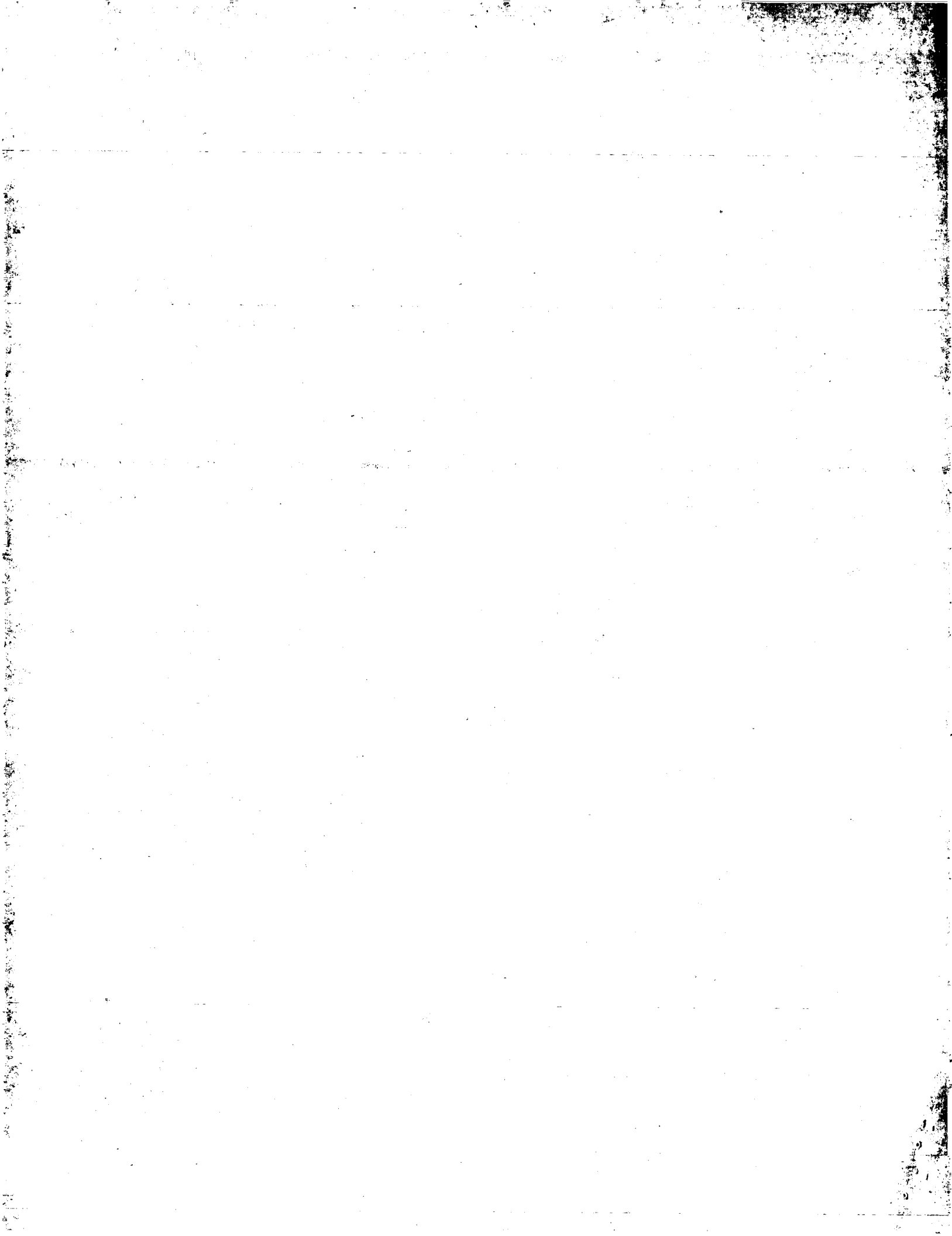
QY 455 agccccagcatcttctccttggcaaaagctcaaacatcctctctgcaagtcacatcttagga 514

Db 396 AGTCCCAATACCTCGCTGTTGGCAAAATTTAAACATAAACTAAAAAATACACTATCAAT 455

QY 515 ttcttcttttctcatgggtcctccaacatgttctcatgtgt 553

Db 456 GCCTTCTCTATATTTGGTCTTTTAATTTGTCCTTCAGT 494

Search completed: July 21, 2002, 02:40:48
Job time: 11698 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 03:56:34 ; Search time 107.09 Seconds
(without alignments)
2555.194 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.4	5.6	7218	1	US-08-232-463-14
2	56.2	5.0	8353	3	US-08-611-587-1
3	48.4	4.3	246240	2	US-08-724-394A-20
4	48.4	4.3	246240	2	US-08-724-394A-21
5	48.4	4.3	246240	2	US-08-724-394A-22
6	47.4	4.2	638	4	US-09-328-111-275
7	44.4	4.0	497	4	US-09-328-111-109
8	43.3	3.9	152331	3	US-09-128-155-16
9	40.6	3.6	2270	2	US-08-852-807-16
10	40.6	3.6	13674	2	US-08-852-807-1
11	40.6	3.6	17410	1	US-07-841-646-3
12	40.6	3.6	17410	1	US-08-447-023-3
13	40.6	3.6	17410	1	US-08-447-570-3
14	40.6	3.6	17410	2	US-08-449-700-3
15	40.6	3.6	17410	2	US-08-449-699A-3
16	40.6	3.6	17415	3	US-08-486-343A-1
17	40.6	3.6	17415	5	PCT-US95-07349-1
18	39.3	3.5	504	4	US-09-328-111-120
19	36.2	3.2	80246	4	US-09-078-294-4
20	36.2	3.2	80595	4	US-09-078-294-3
21	35.4	3.2	289	4	US-09-007-005-17
22	35.4	3.2	289	4	US-09-244-796-17
23	35.2	3.2	2358	3	US-09-022-983-1
24	35.2	3.2	2360	4	US-09-490-692-10
25	35.2	3.2	40352	3	US-08-846-111B-15
26	34.2	3.1	4002	3	US-09-356-952-9
27	33.6	3.0	437	4	US-08-576-202-29

Sequence 29, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14


```
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-20

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1011 aaatgttatggtttaaatattgtctctccaaagctcatgttgaaatttaatacgccaat 1070
||| ||||||| ||||||| ||| ||||||| ||| ||| ||||||| ||| ||| |||
Db 142977 AAGTGATATGTTGGATATTGTCCCTCTAAATCTCATGTGTAATCTCAAT 142918

Qy 1071 gtggca 1076
||| |
Db 142917 GTTGA 142912

RESULT 5
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-22

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1011 aaatgttatggtttaaatattgtctctccaaagctcatgttgaaatttaatacgccaat 1070
||| ||||||| ||||||| ||| ||||||| ||| ||| ||||||| ||| ||| |||
Db 142977 AAGTGATATGTTGGATATTGTCCCTCTAAATCTCATGTGTAATCTCAAT 142918

Qy 1071 gtggca 1076
||| |
Db 142917 GTTGA 142912

RESULT 4
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-21

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
```

Db 142917 GTTGGGA 142912

RESULT 6

US-09-328-111-275/c

; Sequence 275, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/328,111

; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 275

; LENGTH: 638

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(638)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-275

Query Match

Best Local Similarity 4.2%; Score 47; DB 4; Length 638;

Matches 62; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 990 accaaattattcaagactgcgaattgttgaatttgaatttgcctctcccaagctca 1049

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 335 ACAAGTTCCTCTCCACCATCTGATGTTGGATGCTCTCCCTCCAAATCTCA 276

Qy 1050 tgttgaaatttaatcgccaatgtggca 1076

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 275 TGTGAAATGTAATTCCCACTGTTGGA 249

RESULT 7

US-09-328-111-109/c

; Sequence 109, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCD-257 (US)

US-09-328-111-109

; CURRENT APPLICATION NUMBER: US/09/328,111

; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 109

; LENGTH: 497

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(497)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-109

Query Match

Best Local Similarity 4.0%; Score 44.4; DB 4; Length 497;

Matches 54; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1008 tgcaaatgttatggttaaatattgtctctcctcaagctcatgttgaaatttaacgcc 1067

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 319 TCCTACTGATATGTTGGATGTTTGTCCCTCCAAATCTCATGATGAATGTAATCCCC 260

Qy 1068 aatgtggcag 1077

||||| |||

Db 259 AGTGTGGAG 250

RESULT 8

US-09-128-155-16

; Sequence 16, Application US/09128155

; Patent No. 6117654

; GENERAL INFORMATION:

; APPLICANT: Pan, Yang

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY

; FILE REFERENCE: 09404/052001

; CURRENT APPLICATION NUMBER: US/09/128,155

; CURRENT FILING DATE: 1998-08-03

; EARLIER APPLICATION NUMBER: US 60/091,650

; EARLIER FILING DATE: 1998-07-02

; EARLIER APPLICATION NUMBER: US 60/054,646

; EARLIER FILING DATE: 1997-08-04

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 16

; LENGTH: 152331

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(152331)

; OTHER INFORMATION: n = A,T,C or G

US-09-128-155-16

Query Match

Best Local Similarity 3.9%; Score 43; DB 3; Length 152331;

Matches 52; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1012 aatgttatggttttaaatattgtctctcctcaagctcatgttgaaatttaacgccaatg 1071

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 73474 aataatggttgatgtttgccccctccaaatctcacatgtgaatgttaatcccccaatg 73533

Qy 1072 tggcagt 1078

|||||

Db 73534 ttgaagt 73540

RESULT 9

US-08-852-807-16/c

```

: Sequence 16, Application US/08852807
: Patent No. 5861298
: GENERAL INFORMATION:
: APPLICANT: Debouck, Christine
: APPLICANT: Drake, Fred
: APPLICANT: Gowen, Maxine
: APPLICANT: Rood, Julie
: APPLICANT: Hastings, Gregg
: APPLICANT: Adams, Mark
: APPLICANT: Fraser, Claire
: APPLICANT: Lee, No. 5861298man
: APPLICANT: Kirkness, Even
: APPLICANT: Blake, Judith
: APPLICANT: Fitzgerald, Lisa
: TITLE OF INVENTION: CATHEPSIN K GENE
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Smithkline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-2799
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/852,807
: FILING DATE: 07-MAY-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/019,942
: FILING DATE: 14-JUNE-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/020,273
: FILING DATE: 17-JUNE-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/026,273
: FILING DATE: 26-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Han, William T
: REGISTRATION NUMBER: 34,344
: REFERENCE/DOCKET NUMBER: ATG50006-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5219
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2270 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: US-08-852-807-16

```

Oy 1015 gttatggtttaaatattgtctcctcccaagctcatgttgaaatttaacgcccaatggg 1074
+
Df 576 GATATGGTTTGATGTGTCCTCCCTCCCAGTCTCATGCTGAANCTGATCCCCAATGCG 517
+
Oy 1075 cag 1077
++

```

Db      516 GAG 514

RESULT 10
US-08-852-807-1/c
; Sequence 1, Application US/08852807
; Patent No. 5861298
; GENERAL INFORMATION:
; APPLICANT: Debouck, Christine
; APPLICANT: Drake, Fred
; APPLICANT: Gowen, Maxine
; APPLICANT: Rood, Julie
; APPLICANT: Hastings, Gregg
; APPLICANT: Adams, Mark
; APPLICANT: Fraser, Claire
; APPLICANT: Lee, No. 5861298man
; APPLICANT: Kirkness, Ewen
; APPLICANT: Blake, Judith
; APPLICANT: Fitzgerald, Lisa
; TITLE OF INVENTION: CATHEPSIN K GENE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBN Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,807
; FILING DATE: 07-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,942
; FILING DATE: 14-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,273
; FILING DATE: 17-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/026,273
; FILING DATE: 26-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50006-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-852-807-1

```

Query Match 3.6%; Score 40.6; DB 2; Length 13674;
Best Local Similarity 77.8%; Pred. NO. 0.072;
Matches 49; Conservative 0; Mismatches 14; Indels 0;
Qv 1015 qttatggttttaaatattttgtctctccaaagctcatgttgaatttaattcgcgaatq

Db 10826 GATATGTTGGATGTGTGTCCTCCCTCCAGTCTCATGCTGAAATGTGATCCCAATGTGG 10767

Qy 1075 cag 1077

Db 10766 GAG 10764

RESULT 11

US-07-841-646-3
Sequence 3, Application US/07841646
Patent No. 5266683
GENERAL INFORMATION:
APPLICANT: OPPERMAN, HERMANN
APPLICANT: OZKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER TYPE: Floppy disk
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/841,646
FILING DATE: 19920221
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 810,560
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 827,052
FILING DATE: 28-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 660,162
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,988
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,849
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 616,374
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 600,024
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 599,543
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 579,865
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 569,920
FILING DATE: 20-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 483,913
FILING DATE: 22-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 422,613
FILING DATE: 17-OCT-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 315,342
FILING DATE: 23-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 232,630
FILING DATE: 15-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 179,460
FILING DATE: 08-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-001CP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17410 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: 3192..3730
OTHER INFORMATION: /label= EXON-1
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10413..10414
OTHER INFORMATION: /label= GAP-1
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 10696..10891
OTHER INFORMATION: /label= EXON-2
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10960..10961
OTHER INFORMATION: /label= GAP-2
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11059..11211
OTHER INFORMATION: /label= EXON-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5


```

;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
;
; US-07-841-646-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.082;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; Qy 1030 ttgtctctccaaagtctgttgaatttaacgcgaatgtgctgactaagaagtg 1089
; ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGTCCTACAAAACATGTTGTAATTTAATGTCAGTGAACATTATTGAGAGTT 8871
;
; Qy 1090 atg 1092
; |||
; Db 8872 ATG 8874
;
; RESULT 12
; US-08-147-023-3
; Sequence 3, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10636..10891
; OTHER INFORMATION: /label= EXON-2
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; THIS SEQUENCE."
;
```

```

; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /note= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-147-023-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.082;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; QY 1030 ttgtctcctccaaagctcatgttgaaatttaatcgccaaatgtggcagtactaagaagtg 1089
;      ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTGTCCCTCAAAAACATCATGTGTAATTAAATGTCACGTGAACATATTGAGAGGTT 8871
;
; QY 1090 atg 1092
;      |||
; Db 8872 ATG 8874
;
; RESULT 13
; US-08-447-570-3
; Sequence 3, Application US/08447570
; Patent No. 5714589
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,570
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; APPLICATION NUMBER: US 827,052
;
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-447-570-3

Query Match          3.6%; Score 40.6; DB 1; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.082;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1030 ttgtctctccaaagtcattgttgaatttaatcgccaatgtggcagtgactaagaagtg 1089
      ||||| | || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTGTGCTCTACAAAACATCATGTTGAAATTTAATGCCAGTGTAAACATTATTGAGAGGTT 8871

QY 1090 atg 1092
      |||
Db 8872 ATG 8874

RESULT 14
US-08-449-700-3
; Sequence 3, Application US/08449700
; Patent No. 5863758
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTROGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
```

```
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,700
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 03:57:56 ; Search time 376.86 Seconds
(without alignments)
5075.207 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229.8	20.6	1331	18	AAT69546 Rat pheromone rece
2	228	20.5	1538	18	AAT69549 Rat pheromone rece
3	220.2	19.8	1386	18	AAT69545 Rat pheromone rece
4	215	19.3	584	22	ABA60235 Human foetal liver
5	215	19.3	584	22	ABA28536 Probe #7002 for ge
6	215	19.3	584	22	AAK08511 Human brain expres
7	215	19.3	584	22	AAK34394 Human bone marrow
8	215	19.3	584	22	AAI40117 Probe #8803 used t
9	209	18.8	1496	18	AAT69547 Rat pheromone rece

10	207.4	18.6	1053	18	AAT69548 Rat pheromone rece
11	206	18.5	1265	18	AAT69550 Rat pheromone rece
12	186.2	16.7	1313	24	AAS98047 Human DNA for pote
13	183.2	16.4	1015	24	AAD24968 Human G-protein co
14	164.8	14.8	628	22	AAH74901 Probe used to isol
15	156.8	14.1	632	22	AAH74901 Probe used to isol
16	152.4	13.7	461	22	AAK06912 Human brain expres
17	152.4	13.7	461	22	AAK32640 Human bone marrow
18	152.4	13.7	461	22	AAK32640 Probe #7141 used t
19	149.4	13.4	632	22	AAH74899 Probe used to isol
20	147.8	13.3	628	22	AAH74900 Probe used to isol
21	145	13.0	653	22	AAH74903 Human purified sec
22	143.8	12.9	640	22	AAS63212 Human purified sec
23	139.6	12.5	662	22	AAH74902 Probe used to isol
24	137	12.3	636	22	AAH74905 Probe used to isol
25	134.8	12.1	300	22	ABA72774 Human foetal liver
26	134.8	12.1	300	22	ABA38413 Probe #16879 for g
27	134.8	12.1	300	22	AAK21203 Human brain expres
28	134.8	12.1	300	22	AAK47360 Human bone marrow
29	134.8	12.1	300	22	AAI53196 Probe #21882 used
30	133.8	12.0	574	24	AAS98151 Human DNA for pote
31	130.2	11.7	512	24	ABA05725 Rat pheromone rece
32	128.6	11.5	512	24	ABA05712 Rat pheromone rece
33	128.6	11.5	512	24	ABA05731 Rat pheromone rece
34	127	11.4	512	24	ABA05704 Rat pheromone rece
35	127	11.4	512	24	ABA05705 Rat pheromone rece
36	127	11.4	512	24	ABA05710 Rat pheromone rece
37	125.8	11.3	512	24	ABA05721 Rat pheromone rece
38	125.2	11.2	530	18	AAT69551 Human pheromone re
39	122.4	11.0	512	24	ABA05708 Rat pheromone rece
40	122.4	11.0	512	24	ABA05703 Rat pheromone rece
41	120.8	10.8	512	24	ABA05701 Rat pheromone rece
42	120.8	10.8	512	24	ABA05702 Rat pheromone rece
43	120.8	10.8	512	24	ABA05711 Rat pheromone rece
44	120.8	10.8	512	24	ABA05714 Rat pheromone rece
45	120.8	10.8	512	24	ABA05715 Rat pheromone rece

ALIGNMENTS

RESULT 1
AAT69546
ID AAT69546 standard; DNA; 1331 BP.
XX AAT69546;
AC AAT69546;
DT 26-AUG-1997 (first entry)
XX Rat pheromone receptor VN3 cDNA.
DE
XX Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW maternal behaviour; reproductive behaviour; fertility;
KW hormone secretion; ss.
XX
XX Rattus sp.
OS
XX Key Location/Qualifiers
FH CDS 207..1142
FT /*tag= a
FT 1038..1040
FT /*tag= b
FT /*transl_except= pos:1038..1040:_aa:Gly
FT /*note= "GTA codes for valine"
FT 1086..1088
FT /*tag= c
FT /*transl_except= pos:1086..1088:_aa:Glu
FT /*note= "GGA codes for glycine"
XX
XX WO9714790-A1.
XX 24-APR-1997.
XX


```
Db 278 taagaacagcagctctacactgattcttaacataaagaataacacctttttctgctgaaattgg 337
Qy 178 ctttggattttcagaacacacactctctctcttttttccacatcttccactttgttttcag 237
Db 338 cattggagcttcagccaatagctccctactctctcttccacatcttcaagttaatttgg 397
Qy 238 tcacaggttaagtccattgacatgataaattagtcacactgtctctctccacataactgct 297
Db 398 gcagagttccagactcactgacctgccattgtctcttctccataactcaactactat 457
Qy 298 gctcttcaactcaggcaatattgtgtcttagactcttcttgggtccacagaataactcagga 357
Db 458 gctactgatgagcgc---attcatagccacagacacttttatttcttggagaggggtggga 514
Qy 358 tgacttaggtataagtgattgtctttttaacaaggtgatgaggggacctctccactg 417
Db 515 tgacatcatatgtaaacctctctctacctgtacagaaacttttagaggtctctctcttgg 574
Qy 418 caccctctgctctgagtggtctcaggccataca---tcagccccagcatctctctcctt 474
Db 575 taccagctgctgttgagtgctcggagccatctcactcagtcaccagagctcctgttt 634
Qy 475 ggcagaagctcaaacatctctcgaagtcacatcttaggattctcttcttctcatgggt 534
Db 635 agcaagttcaaacataaagccttcccatcactctctctgtgccattcttctcttgagctg 694
Qy 535 cctcaacatcttcatt-ggtgtaactctctgtgtacactgcggctacccccagtgaaac 593
Db 695 cctctacatgttctattagcagtcacactcttagtaccatcatgtgcccccccaatttgac 754
Qy 594 ggggcccagtc-ttctgtttgtacacagcactgtctcttcttggccatgagctacaccca 652
Db 755 caagaaatgacttattcatgttactcagtggtgctcttcttctacccatgattacctcat 814
Qy 653 caggagactgttttccacactaatgactttgagggatgtcacctttatagggttcaggt 712
Db 815 gcaagcatgttttctacactgtggccatcagggatgtcttcttattagttcctcaggt 874
Qy 713 cctctcaagaggctacatgttgattattttatcacagacaataagaggctatctcagtgcc 772
Db 875 cctgtcaacatggtacatggtggctctctgtgtagtc-acaggaaacagaccggcgtac 933
Qy 773 ttccacagccagcgtctccccc--gagctcacagtgaaagagcgtccaggctatct 830
Db 934 ttcagggtaccagcctttcccccacaaagcatcccccagacaaggggccacccttccatcc 993
Qy 831 tactgtggtgaggtttgtcttccacactggtgggtgacttaogttctctatcttttcaggag 890
Db 994 tgatgctcagcgttattgttctgtatgtctgtcttggacagcatgtgtcgcagctcaa 1053
Qy 891 ggtgtacatggataaattgattctctgtctgtgtgtggtccagggtattgtggccaatagct 950
Db 1054 gaactatgtatctgaagatcccaatcttattcttatacaactatttattgtgcaactct 1113
Qy 951 atgcccgaattagctctttagctcaattatgtctgatacccaataattcaagaactctg 1009
Db 1114 atgcccagtaagccctttgtgttattgtcactgaaaaacatatagttaactctttg 1172

RESULT 4
ABA60235 standard; DNA; 584 BP.
XX
AC ABA60235;
XX
DT ABA60235;
XX
DE 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8540.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
```

```
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 8540; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fip.wipo.int/pub/published\_pct\_sequences.
XX
SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.3%; Score 215; DB 22; Length 584;
Best Local Similarity 64.7%; Pred. No. 2.7e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;

Qy 190 agcaaacacacatccttcttcttccacatcttccactttgttttccagtcacaggtctaa 249
Db 4 agcaaacacatctctcttcttccaaatcttctcactctctcaggtacagggcctaa 63
Qy 250 gtccattgacatgataattagtcacctgtctctcatcacatactgtctcttccactca 309
Db 64 gccactgacccatacactgtcacctggccctgtccacttaggaggtctctcattgt 123
Qy 310 ggcaatattggttccttagactcttgggttcacagaatactcaggtatgattaggtta 369
Db 124 ggtctctggcatctccagacctgtttgagtcactgtatttccagaatgacttcaagt 183
Qy 370 taaggctcattgtcttttaaacaaaggtgatgaggggctctccatctgcacccctcct 429
Db 184 taaggcattctctgcacaggggtgatgagggctctcctcctcctgacaccactgct 243
Qy 430 cctgagtgctgcaggc---catcatcagccccagcatcttctccttggcaagctcaa 486
Db 244 cctgagcatgtctccaggctgtgccatcagccctggcacctctctggtctcgaaggat 303
Qy 487 acatctcttgcaggtcacatcttaggattctcttcttcttcttcttcttcttcttctt 545
Db 304 acagaaattcaagggtacatcttctcactctcttcttcttcttcttcttcttcttctt 363
Qy 546 tcattgggttaattctgtgtacactgcgggtacccccagtgaaacgggggcagctctt 605
Db 364 tctcagtagtaactgctctctccactgtggtcttcttcttaattggggacgaagctgt 423
Qy 606 -ctgtttgtacacagcactgttcttcttggccatgagctacacccacagagactgtt 664
```

Db 424 gctaagtacagtaataactgatcacattttctccataaagctacattatcaggagcctgtc 483
Qy 565 ttccacataatgactttgagggatgacacttttatagggttcattgtctctcctaagagg 724
Db 484 tttcatgtcgccactgcttacgaatgtctctttgttagcaatcatcgagccctcaagtgc 543
Qy 725 ctacatggtgattattttatcacagaca 751
Db 544 atacatggttaattctcttgttcaggca 570

RESULT 5
ABA28536
ID ABA28536 standard; DNA; 584 BP.
XX
AC ABA28536;
DT 23-JAN-2002 (first entry)
XX
DE Probe #7002 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1; SEQ ID No 7002; 530pp; English.
XX
PS The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

Query Match 19.38; Score 215; DB 22; Length 584;
Best Local Similarity 64.78; Pred. No. 2.7e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;
Qy 190 agcaaacaccatctctttttccacattcttcacattgttttttcagtcacaggtctaa 249

Db 4 agcaaacacttctctctatttctccaaatcttctcactcttcaggatacacaggcttaa 63
Qy 250 gtccattgacatgataattagtcacctgtctctcaccatactctgtctctcactca 309
Db 64 gccacagaccatccactgtccactggccctgtgtccacttaggtatgctccctattgt 123
Qy 310 ggcaattatggtgctccttagactcttctgttcacagaatactcaggatgatcttaggta 369
Db 124 ggtcttctgtcctcagacacctgttggagtcactgtattttcagaatgactccaagt 183
Qy 370 taaggtcatigtcttttaaacagggtgatggagggcctctcccatctgcacccctgct 429
Db 184 taaggcatcttctgtcatgacaggggtgatggagggcctctcccatctgcacccctgct 243
Qy 430 cctgagtggtctccaggc---catcatcagcccccagcatcttctccttggccaaagctcaa 486
Db 244 cctgagcatgtccaggctgtgcgcacatcagccctggcacctcctggtctgcaaggattaa 303
Qy 487 acatcctctgcgaagtcatcatcttaggattctcttcttctcattggtctcctcaacatgt- 545
Db 304 acagaaattcaaggttacatcttctcactcctcttctctctctggttctcagttgtc 363
Qy 546 tcattggttaattctctgtctgacactggcgtcaccgccagtgaaacggggccagtcct 605
Db 364 tctcagtagtaacctgtctctcactgtggtcttcttcttaattgggacccaagcgttgt 423
Qy 606 -ctgtttgtcatcacgacactgttctctttttgtcccatgagctacacccacaggagactgtt 664
Db 424 gctaaagtatcagtaataactgatcacatttcttctccataagctacattatcaggagcctgtc 483
Qy 565 ttccacataatgactttgagggatgacacacttttatagggttcattgtctctcctaagagg 724
Db 484 tttcatgtcgccactgcttacgaatgtctctttttgttagcaatcatcgagccctcaagtgc 543
Qy 725 ctacatggtgattattttatcacagaca 751
Db 544 atacatggttaattctcttgttcaggca 570

RESULT 6
AAK08511
ID AAK08511 standard; DNA; 584 BP.
XX
AC AAK08511;
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 8502.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;

QY 606 -ctgtttgtatcacagcactgttctcttctttgttcccatagagctacacccacagagagactgtt 664
 Db 424 gctaagtatcagtaaatcactgatcactcttcttccatagctacatcattatcaggagcctgtc 483
 QY 665 ttccacactaatgactcttgaggagtgacactttataggggttcattgtctctctcaagagg 724
 Db 484 ttctcatctgccactgcttacgaatgtctcttcttctgttagcaatcattcagccctcaagtgc 543
 QY 725 ctacatggtgattattttatcacagaca 751
 Db 544 atacatggttaattctctgttcaggca 570

RESULT 8
 AAI40117
 ID AAI40117 standard; DNA; 584 BP.
 XX
 AC AAI40117;
 DT 17-OCT-2001 (first entry)
 DE Probe #8803 used to measure gene expression in human placenta sample.
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 OS Homo sapiens.
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234887.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX Claim 25; SEQ ID No 8803; 654pp; English.
 XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.3%; Score 215; DB 22; Length 584;
 Best Local Similarity 64.1%; Pred. No. 2.7e-50;
 Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;

QY 190 agcaaacacacactctcttctttccacattctcactttgttttcagtcacagggtctaa 249
 Db 4 agcaaacacactctctctattcttccaaattctcactctcttcaggtacacagggtctaa 63
 QY 250 gtccattgacatgataattgacactgtctctctcattccacatactgctgtcttcaactca 309
 Db 64 gccacatgaccacatcacctgtcacctggcccttgctccacttagggatgctctcattgt 123

QY 310 ggaatatattgtgtctcttagactcttcttgggttccagaataactcagagatgatcttaggta 369
 Db 124 ggtctcttggcatctccagacacctgtttgagtcactgtattttcagaatgacttcaagtg 183
 QY 370 taaggctcatctctttttaacaagggtgatgaggggcctctcctcactgcacccctgcct 429
 Db 184 taaggcattctctgcattgcacaggggtgatgaggagcctctcctcactgcacccctgcct 243
 QY 430 cctgagtggtctccaggc---catcatcagcccccagcatctctccttggcacaagctcaa 486
 Db 244 cctgagcatgctccaggcgtctgcacatcagccctggcacctcctgggtctcagaagattaa 303
 QY 487 acatcctcttgcaggatcacatttagattcttctcttcttctcattggtctcacaacatgt- 545
 Db 304 acagaaattccaagggttacatcttctcactctctcttctctcttctcctcagttgtgc 363
 QY 546 tcattggttaattctctgtgtacactgggctaccccccagtgaaacggggccagctctt 605
 Db 364 tctcagtagtaacctgctctcctcactgtggtcttcttcttaattgggacccaagacggttgt 423
 QY 606 -ctgtttgtcatcacagcactgttctcttcttggcccatgagctacacccacagagactgtt 664
 Db 424 gctaaagtatcagtaaatcactgatcactttcttccataagctacattatcaggagcctgtc 483
 QY 665 ttccacactaatgactttgaggagtgatgcacacttttatagggttctcctcctcacaagagg 724
 Db 484 ttctcatgctccactgcttacgaatgtcttcttcttcttcttctgtagcaatcagcagcctcaagtgc 543
 QY 725 ctacatggtgattattttatcacagaca 751
 Db 544 atacatggttaattctctgttcaggca 570

RESULT 9
 AAT69547
 ID AAT69547 standard; DNA; 1496 BP.
 XX
 AC AAT69547;
 DT 26-AUG-1997 (first entry)
 XX
 DE Rat pheromone receptor VN4 cDNA.
 KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;
 KW maternal behaviour; reproductive behaviour; fertility;
 KW hormone secretion; ss.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 CDS 351..1283
 FT /*tag= a
 FT
 XX WO9714790-A1.
 XX
 XX
 PD 24-APR-1997.
 XX
 PF 18-OCT-1996; 95WO-US16637.
 XX
 PR 19-OCT-1995; 95US-0005698.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Axel R, Dulac C;
 XX
 DR WPI; 1997-245107/22.
 DR P-PSDB; AAW19105.
 XX
 PT Nucleic acid molecule encoding vertebrate pheromone receptor -
 PT useful to identify modulators for control of reproductive and social
 PT behaviour, fertility and hormone secretion
 XX


```
Matches 471; Conservative 0; Mismatches 321; Indels 11; Gaps 6;
QY 207 tcttttccacatcttccacttctgttttccagtcacaggtctcaagtcacattgacatgataa 266
Db 1 tttttccacactctctcgtcttctgaaagaaacagatctcaagccattgatctctaca 60
QY 267 ttagtcaactgtctctcaccacatactcgtctcttcaactcaggcaataattggtgctc 326
Db 61 tgcgtctcttcccttaaccacaaactcaagtgctctataactattggacttatagctgc-- 118
QY 327 tagactcttcttggttcacagaatactcagtgatctcttaggtataagggtcattgtctttt 386
Db 119 -agacatgtttatgctcggggagatgggattctaccacatgccagtcctctattctatt 177
QY 387 taacaagggtgtagggggcctctccatctcgaacccccctgcctcctgtagtgctccagg 446
Db 178 tggacaggcttttgagggttttaccttctgtctacgtctgtctgaatgctctttgga 237
QY 447 ccataca---tcagccccagcatctctctctggcaagctcaaacatcctctcgaagtc 503
Db 238 ccatacctcagtcctcagaagctcctgttttaaacacatttaaacataaatctccccatc 297
QY 504 acattttaggtattctcttctctcagtggtctcacaatggtctcattgg-tgtaattcttc 562
Db 298 acatctcaggtgccttcttcttctgtgtctctatatattctttggcagtcacctct 357
QY 563 tgcctacactcggctacccccccagtgaaac-ggggccagtccttctgttctgctacagc 621
Db 358 ttttatacaaatgtctacccccctttagtctcagataattttatgtatgattactaaat 417
QY 622 actgttctcttttgcctcagtcacccacacagcagagactgtttttcacactaatgactt 681
Db 418 cctgttcttctacccatgagttaactcagaaacagcatgttttccacaccaaaggcca 477
QY 682 tgagggtatgccttcttataggtgttcattggtcctcctcaagaggtcacatggtgattatt 741
Db 478 tcagggaagccttcttattgtctctcattggtcctgtcctgtccagtggtacatggtgttcttc 537
QY 742 tatacagacaataagagctatctcagtcgcttcacacagcagcctgtc--cccgagtc 799
Db 538 tatggagac-acaaatacaggcccgccgcatcttcacagcaccagccttcttccaaaagt 596
QY 800 tcaccagtgaaagagcctccacagctatcttactcgtcgtgagttgttcttcacatac 859
Db 597 tccccagcaagagggccaccagcaccatcatgattctcatgacttcttctgtgtcttc 656
QY 860 tgggtggaactttacgtctctcatttccagggaggtgacatgataaagtctctctgtc 919
Db 657 tacattttggaaaattgtgtcttctactctaggtgacattcaaggatgggtcaatgttc 716
QY 920 gtgtggtccaggttatgttgcccaatagctatgcgcaattagtccttctgtatgctaatt 979
Db 717 tactgtgtccaaattatgtgtgtcccatagctatgcccacatcagccttcttctgtttatt 776
QY 980 tatgtgtataaccaaatattcaa 1002
Db 777 tgcacagaaaagcgtataattaa 799
RESULT 11
AAT69550
ID AAT69550 standard; DNA; 1265 BP.
XX
AC AAT69550;
XX
XX 26-AUG-1997 (first entry)
DT
DE
DE
DE
XX
XX Rat pheromone receptor VN7 cDNA.
KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW maternal behaviour; reproductive behaviour; fertility;
KW hormone secretion; ss.
XX
```

```
OS Rattus sp.
XX
FH Key Location/Qualifiers
CDS 138..1061
FT /*tag= a
FT
XX
XX WO9714790-A1.
PN
XX
XX 24-APR-1997.
PD
XX
XX 18-OCT-1996; 96WO-US16637.
PF
XX
XX 19-OCT-1995; 95US-0005698.
PR
XX
XX (UNCO ) UNIV COLUMBIA NEW YORK.
PA
XX
XX Axel R, Dulac C;
PI
XX
XX WPI; 1997-245107/22.
DR
XX
XX P-PSDB; AAM19108.
XX
XX Nucleic acid molecule encoding vertebrate pheromone receptor -
PT useful to identify modulators for control of reproductive and social
PT behaviour, fertility and hormone secretion
XX
XX Claim 1; Fig 13; 123pp; English.
PS
XX
XX cDNA clones (AAT69546-50) respectively code for rat pheromone
CC receptors VN3, VN4, VN5, VN6 and VN7 (AAM19104-08), members of a
CC novel family of presumed 7-transmembrane domain receptors that are
CC evolutionary independent of the odorant receptors of the main
CC olfactory epithelium. These clones, and a clone for VN2 (sequence
CC not provided), were isolated from rat vomeronasal organ cDNA
CC libraries by PCR and hybridisation. A differential cloning method
CC was used to isolate VN7 cDNA (AAT69545). VN1-VN7 (not VN2) clones
CC have been deposited in pBluescript as ATCC 97294-97299. A human
CC homologue, HG25 (AAT69551), has also been isolated. VN polypeptides
CC can be expressed in host-vector systems for use in identifying
CC modulators for control of maternal, reproductive and social
CC behaviour, to increase fertility, control hormone secretion and to
CC regulate food uptake in humans and animals.
XX
SQ Sequence 1265 BP; 325 A; 287 C; 226 G; 426 T; 1 other;
Query Match 18.5%; Score 206; DB 18; Length 1265;
Best Local Similarity 55.7%; Pred. NO. 1.3e-47;
Matches 498; Conservative 0; Mismatches 385; Indels 11; Gaps 5;
QY 122 atgatcaaaactcctcgtttttattaccatacaaaattcttctttatccacaagccagcttt 181
Db 138 atgatgaacctgttctatggctccagatgacaaacatgatctcatatcagggaactgtt 197
QY 182 ggaatttcagcaaaaacaccatccttcttttccacatcttccacctttgtttttcagtcac 241
Db 198 cgcaccttccctaaacagtatcctgtttttggccacctcgtcatgttctttgaagaaac 257
QY 242 aggttaagtcacattgacatgataaattagtcacctgtctctcacaatactcagtcgtc 301
Db 258 aggttaagcccatgtatcgtgtgcatcttcttcttcttcccttaaccacaataatgctgt 317
QY 302 ttcactcaggcaaatattggtgccttagactcttcttggttcacagaataactcagatgat 361
Db 318 gtaactatgggaactcatag---ctgcagacatggttatggctcaggggatatcgggatatt 374
QY 362 cttaggtataaggtcatgtctttttaacaagggtgtagggggcctctccctatctgcacc 421
Db 375 accacatgcagggtcccttatctattttcacaagacttttgagggggtttcaacctttgtgct 434
QY 422 cctgtcctctgagtgctccagggccatca---tcagcccagcatcttctctcttgga 478
Db 435 gcctgtctactgcatactcctttggaccttcaactcactcagtcctcagtcctcagtcctcgtttaaca 494
```


Db 366 tgtctttcagtagtacataattttctcactgtggtcttcttcattgtgacccagacca 425
Qy 601 gtctctgtgtgtacacagcactgttctcttcttttggccatgagtagtacaccacagggagac 660
Db 426 atctacttaaggtccgcaaatcactgctcactgcttctcccatgaaatccatcatgtggggag 485
Qy 661 tgtttttcacactaagtactgttgaggatgtgcacatttataggtttcattggtctctcaaa 720
Db 486 tgttttcc-----ttgtaggattacgtgctctcaaa 516
Qy 721 gaggtcacatggtgattattttatcacagacaataagaggctatctcagtgcttcacaca 780
Db 517 gtgcatacatgatgatctttttgtcca-agcatcagaagtgtcccgacatcttcacagt 575
Qy 781 gcaagcctgtcccc--gagttccaccagtgaaagagagcctcccaggtctatcttactcgtg 838
Db 576 accagcctttccccaagatctctccagagaaaagggttacccagatcatctgcccactg 635
Qy 839 gtgagt---tttgtcttcacatactgggtgagcttttacgttctctcatitttcaggaggtgtg 895
Db 636 gtgaattgctttgttcagtgttctgggtgagaccttatcatctcatctcttcttccctg 695
Qy 896 acatggtataaatgattctctgtcgtagtgtggtccaggttatttggccaatagctatgcc 955
Db 696 ttatggagctataaccagtcactcctgagcatctagaacctgtgtgctgtctatgcc 755
Qy 956 gcaattagctcttgatgctaatttatgtcgtataaccaaataatcaagactctgcaaatg 1015
Db 756 actcgtcttccattggtacaaaatccgtctcgtgataaaagatgcaatattctccaaaaa 815
Qy 1016 ttatggttttaaat 1028
Db 816 atggaattaaagt 828

RESULT 13

AAD24968
ID AAD24968 standard; cDNA; 1015 BP.
AC AAD24968;
XX
XX
XX 12-MAR-2002 (first entry)
DE Human G-protein coupled receptor-13 (GCREC-13) cDNA.
KW Human; G-protein coupled receptor-13; GCREC-13; therapy; cancer; stroke;
KW cell proliferative disorder; neurological; epilepsy; Parkinson's disease;
KW Alzheimer's disease; inflammation; thyroiditis; haemolytic anaemia; AIDS;
KW Acquired Immune Deficiency Syndrome; dementia; nontropic; cholelithiasis;
KW multiple sclerosis; atherosclerosis; angina pectoris; gastroenteritis;
KW diabetes; ulcer; viral infection; immunosuppressive; ss.
XX Homo sapiens.
OS
FH
FH Key Location/Qualifiers
FT CDS 248..1015
FT /*tag= a
FT /product= "Human GCREC-13 protein"
XX
XX WO200198351-A2.
XX
XX 27-DEC-2001.
XX
XX 15-JUN-2001; 2001WO-US19275.
PF
XX
XX 16-JUN-2000; 2000US-212483P.
PR 22-JUN-2000; 2000US-213954P.
PR 29-JUN-2000; 2000US-215209P.
PR 07-JUL-2000; 2000US-216595P.
PR 14-JUL-2000; 2000US-218936P.
PR 19-JUL-2000; 2000US-219154P.
PR 21-JUL-2000; 2000US-220141P.
XX

(INCY-) INCYTE GENOMICS INC.

PA Lal P, Baughn MR, Hafalia AJA, Nguyen DB, Gandhi AR, Kallick DA;
XX Griffin JA, Yue H, Khan FA, Patterson C, Lu DAM, Tribouley CM;
PI Lu Y, Walia NK, Graul R, Yao MG, Yang J, Ramkumar J, Au-Young J;
PI Elliott VS, Hernandez R, Walsh RT, Borowsky ML, Thornton M, He A;
XX WPI; 2002-075627/10.
DR P-PSDB; AAE15643.

Isolated human G-protein coupled receptor polypeptides and the use of
these sequences in the diagnosis, treatment and prevention of diseases
and in the assessment of exogenous compounds on the expression of the
receptors -

Claim 11; Page 139; 143pp; English.

The invention relates to isolated human G-protein coupled receptor
(GCREC) polypeptides and their biologically active fragments. GCREC and
protein is useful in treating a disease or condition associated with an
increase or decrease in expression of functional GCREC. The GCREC's are
useful in the diagnosis, treatment and prevention of cell proliferative
disorders (cancer, leukaemia, melanoma); neurological disorders (stroke,
epilepsy, Parkinson's disease, dementia; Alzheimer's disease); autoimmune
inflammatory disorder (thyroiditis, haemolytic anaemia, AIDS, multiple
sclerosis); cardiovascular disorder (atherosclerosis, angina pectoris),
gastrointestinal disorder (ulcer, cholelithiasis, gastroenteritis),
metabolic disorders (diabetes); viral infections (herpes virus) and in
the assessment of the effects of exogenous compounds on the expression
of the nucleic acid and amino acid sequences. The present sequence is
human GCREC-13 cDNA.

Sequence 1015 BP; 261 A; 226 C; 182 G; 346 T; 0 other;

Query Match 16.4%; Score 183.2; DB 24; Length 1015;
Best Local Similarity 59.1%; Pred. No. 3e-41;
Matches 469; Conservative 0; Mismatches 298; Indels 27; Gaps 8;

Qy 130 acttctcgtgttttattacacataaaaattttttatccacaagcagcttggaaatttc 189
Db 222 actttgcacttgtgtcccatgaaaaatgctcttcatttcttaagctggtattggcatctc 281
Qy 190 agcaaacacacatctcttcttttccacatcttcacctttg-----ttttcagtcaca 242
Db 282 agcaaacacacttctctcttcttcttctgtcgtctctctcttcttcttcttcttgatctca 341
Qy 243 ggttaagtccattgacatgataattagtcacctgtctctcatcacatactgtctct 302
Db 342 ggcagagcgcacttaccctaccagctgtcgtcagtgtggcctcatccacatggtggtcttc 401
Qy 303 tcactcaggccaatattggtgtccttagacttcttgggttcacagaataactcagatgactc 362
Db 402 tcaccatgg---tgttctgtctccacagctcttgaatcactgaatttccagaatgact 458
Qy 363 ttaggataaggtcattgtcttttttaacaaggtgtaggggctctccatctgcacccc 422
Db 459 tcaaatatgaggcatctcttctacacctgaggagggtgtagcaggagaccttccatttgcacca 518
Qy 423 cctgcctcctgagtgtgtctccag--gccatcatcacccccagcatcttctccttgccaa 479
Db 519 cctgcctcctcctggtggtcaggtcgtcacaacatcacccccagcatcttctccttggtgga 578
Qy 480 agctcaaacacatcttctgcaagtccacatcttaggattcttcttcttctcagtggtctccta 539
Db 579 ggtttaaatggaaatccacaatttttacccttccattgttctctcatgtctc-----tctcag 633
Qy 540 acatgttcaattggtgtaattcttctgtctgtacactgcggctaccocccagtgaaacggggcc 599
Db 634 ttttctctgttagtagtgcctgactctttttacactgtggtcttcttcccaatgtgacccag-- 691
Qy 600 agtcttctgttctcatcacagcactgttcttcttttgcctatgactcaccaccagcagaga 659

Db 692 atcaattgcatgctagtaataactgttcaacttttcccaataaaactccataatcagagga 751
Qy 660 ctgtttttcacataaagctttgagggatgtcaccttttatagggttcaggtcctctca 719
Db 752 ctgtttttcactctgtcattatccagagatgttttttaacaaataatgctgtctca 811
Qy 720 agaggctacatggtgattatttatcacagacaataaagaggtctatctcagtgccctcacac 779
Db 812 agtgtcacatgatctcattcaggaactacagag--atcctggtaacctcaag 868
Qy 780 agccagcctgtcccgaggtctcaccagtgaaaagag-cctcccaggtctatcttaactgctg 838
Db 869 cccagcctctacctaaggtctttgcagaggaagccatcagcacatcctgctgcg 928
Qy 839 gtgagttt---tgtctcacatactggtgactttacgttctcattttcagaggtgtg 895
Db 929 gtgagtttctcgggtggaagtacaaagtggacttcaatcattcaacacctcacaagttg 988
Qy 896 acatggataaatga 909
Db 989 ccatgggcatatga 1002

RESULT 14
AAH74904
ID AAH74904 standard; DNA; 628 BP.
AC AAH74904;
XX
DT 29-OCT-2001 (first entry)
XX
DE Probe used to isolate human VNO cDNA sequences.
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
PR 17-FEB-2000; 2000US-0183128.
XX
PA (PHER-) PHERIN PHARM INC.
XX
PI Herman RC, Berliner D;
XX
DR WPI; 2001-522609/57.
XX
PT Novel vomeronasal organ cDNA library useful for identification and
PT isolation of genes expressed in vomeronasal organs, specifically genes
PT for pheromone receptors and ion channels -
XX
PS Example 9; Page 54-55; 59pp; English.
XX
CC The present sequence represents a probe, used to isolate human
CC vomeronasal organ (VNO) cDNA sequences. The cDNA library was
CC constructed from female tissue. Sequences isolated from the VNO cDNA
CC library are useful as a unique resource for the identification and
CC isolation of genes expressed in VNO, specifically genes for pheromone
CC receptors, ion channels and prospective reagents for high throughput
CC assays. They are useful as an excellent source to search for novel
CC genes, gene fragments or other nucleotide sequences encoding proteins
CC that are implicated in detection of pheromones or other vomeropherins
CC in the human VNO. The cDNAs are useful as hybridization probes for
CC determining the presence or concentration of an oligo- or polynucleotide,
CC e.g. DNA, of interest.
XX
SQ Sequence 628 BP; 138 A; 178 C; 123 G; 189 T; 0 other;

Query Match 14.8%; Score 164.8; DB 22; Length 628;
Best Local Similarity 60.3%; Pred. No. 3 6e-36;
Matches 381; Conservative 0; Mismatches 237; Indels 14; Gaps 6;
Qy 258 acatgataattagtcacctgtctctcatccacatactgtctctcactcagggcaatat 317
Db 1 acctgatcatcagctctctgtggccctcatccaccttggatgc---taacagtcattggat 57
Qy 318 tgggtccttagactctcttgggttcacagaataactcaggtatgtcttaggtataaaggtea 377
Db 58 tcagagctgttgatatttttgcatactcagaatgtgtggaatgacatacaaatgcaatccc 117
Qy 378 ttgtcttttaacaacaggtgatgagggcctctccactgcacccctgcctcctcctcagtg 437
Db 118 ttgccacttacacagacttttggagggcctctctcttgtgtacctgtctgtcagta 177
Qy 438 tgttcaggccatca---tcagcccaagcatctctctcttggcgaagctcaaacatcctt 494
Db 178 tcttcaggccatcaaccttagcccaagctctgttagcaaaagtccaataataat 237
Qy 495 ctgcaagtcacatcttaggattctctcttctcatgggtcctcacaacatgttcattggtg 554
Db 238 ccacacagcacagcctgtgtccctctctgtctctgtggcctctctacatctcctgtgta 297
Qy 555 taatcttctgtgtac--actgcgggtaccccccagtgaaacggggcagctctctgtttg 612
Db 298 ctcaactctcttcacacctgtgtgtgactacaactctcttcacgcagctcatattg 357
Qy 613 tcatacagcactgtctctcttttgcctatgagctacacccacagagagactgttttcacac 672
Db 358 tcaatgaatcctgcattatttttaccatggtattacacaccagggatttttttccatat 417
Qy 673 taatgactttgagggatgtcacctttatagggttcaggtctcctcacaagaggtcacatgg 732
Db 418 tggggatatttcgggatgtgtccttcatagggtcctcagccctcctccagcgggtacatgg 477
Qy 733 tgattattttacagacaataaagaggtctatctcagtccttcacacagccagcgtctcc 792
Db 478 tggccctctgtgcagaca-caggaaacagggccagcatcttcacagggatccgctttct 536
Qy 793 cc--gaggtcacagtgaaaagagcctccaggtctctcaggtattactgtggtgag--tttt 847
Db 537 ccaaaagcatccccagagcaaggccaccagggaccatcctgtgtcctcatgagctcttt 596
Qy 848 gtcttcacatactgggtggactttacgttctc 879
Db 597 gtgtgagtactgtgtggactgcaccatatac 628

RESULT 15
AAH74901
ID AAH74901 standard; DNA; 632 BP.
XX
AC AAH74901;
XX
DT 29-OCT-2001 (first entry)
XX
DE Probe used to isolate human VNO cDNA sequences.
XX
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
PR 17-FEB-2000; 2000US-0183128.
XX

PA (PHER-) PHERIN PHARM INC.

PI Herman RC, Berliner D:

WPI: 2001-522609/57.

Novel vomeronasal organ cDNA library useful for identification and isolation of genes expressed in vomeronasal organs, specifically genes for pheromone receptors and ion channels -

PS Example 9; Page 52-53; 59pp; English:

The present sequence represents a probe, used to isolate human vomeronasal organ (VNO) cDNA sequences. The cDNA library was constructed from female tissue. Sequences isolated from the VNO cDNA library are useful as a unique resource for the identification and isolation of genes expressed in VNO, specifically genes for pheromone receptors, ion channels and prospective reagents for high throughput assays. They are useful as an excellent source to search for novel genes, gene fragments or other nucleotide sequences encoding proteins that are implicated in detection of pheromones or other vomerophorins in the human VNO. The cDNAs are useful as hybridization probes for determining the presence or concentration of an oligo- or polynucleotide e.g. DNA, of interest.

Sequence 632 BP: 150 A: 177 C: 109 G: 196 T: 0 other:

Query Match	14.1%	Score 156.8;	DB 22;	Length 632;
Best Local Similarity	59.5%;	Pred. No. 6.4e-34;		
Matches 357; Conservative	0;	Mismatches 232;	Indels 11;	Gaps 5;

Qy 261 tgataattagtcacctgtctctcatccacatactgctgctctctcactcagqcaatatattq 320

Db 1 tggccattgggtctcttgccttaatccaccctactgatactactgatactgagggc---attca 57

Qy 321 tqtccttagactctcttttqqttcacagaatactcagqatgattcttagqataaaggtcattg 380

Db 58 taqccataqacatttttatttcttqqagqqqatqqqatgacatcatatataaatttccttg 117

Qy 381 tcttttaaacaaqqtgatgagggqccctctccatctgcacccccctgcctcctgaqtatc 440

Db 118 tctacttgtacagaaqttttaqaqqctctctctcttqtaccacctqcatqttqaqatqtc 177

Qy 441 tccaqqccatca--tcaqcccccaqcatcttctccttqqcaaaqctcaaacatcccttctq 497

Db 178 tgcaggccatcacccctcagccccagaaqctcctgttttaqcaaaqttcaaacataaqtctc 237

Qy 498 caaqtacacatcttaqqatcttccctttctcatqqgtccctcaacatgttcattggttaa 557

Db 238 cccatcacgtctcctgtgccattatttcqctgagcatcctctacatgtttcattagcaqtc 297

Qy 558 tcttctgctgtacact--gcqgctaccccccaqtgaaacqgqcccaqtcttctgtatttqtca 615

Db 298 acctcttaqtatcccatcaatqccaccccccaatttqaccacgaaacaacttttatqcaagtta 357

Qy 616 tacagcactgttcctttttgcccattgaqctacacccacaggaqactgtttttcacactaa 675

Db 358 ctcaqtccctacattatacccttgaqgttacctcatgcaaaagcatgtttttctacacttc 417

676 **tgactttgagggaatgtcaccttttatagggttcattggtcctctcaaggaggtacatggtga** 735

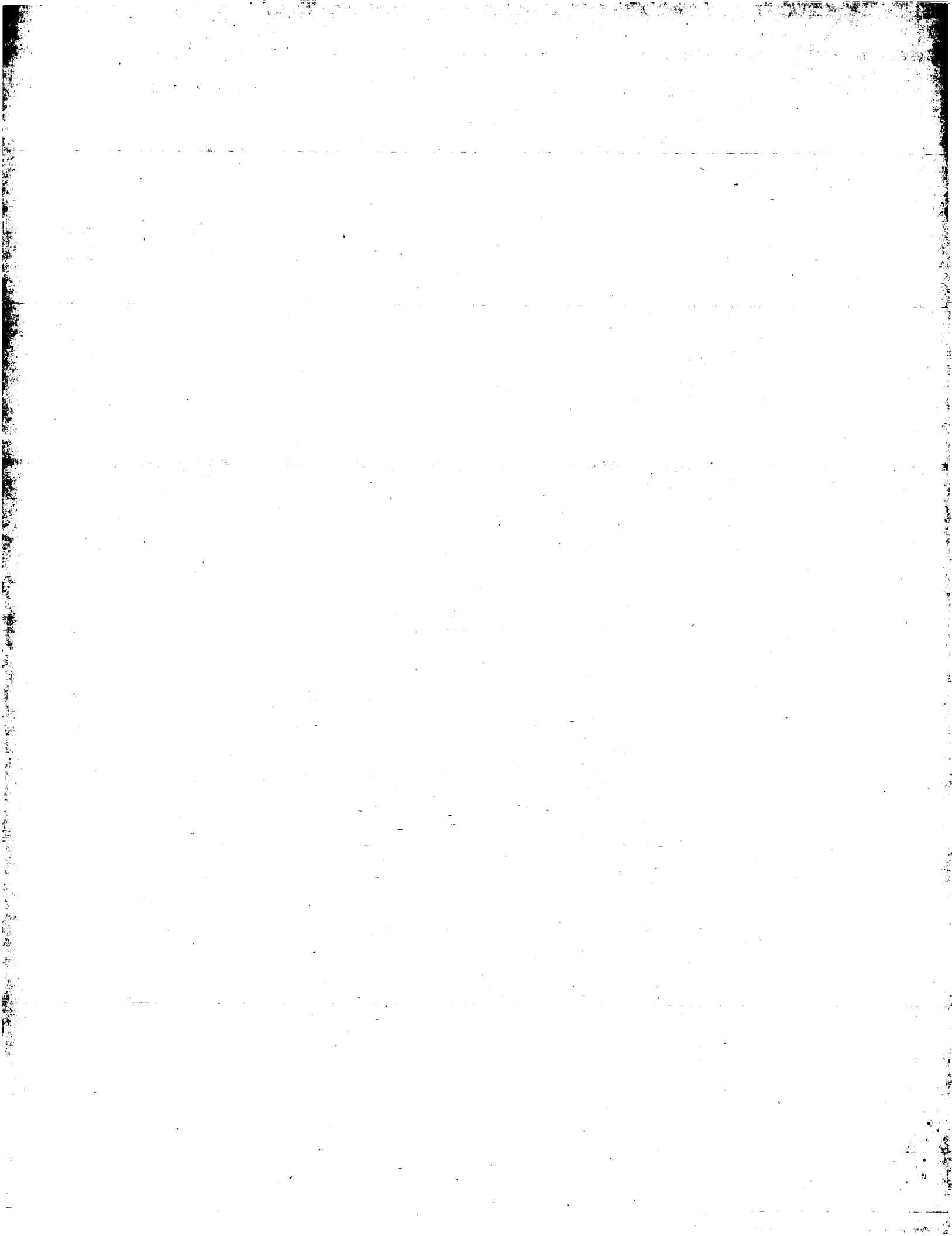
Db 418 tggccatcagagatatctctcttattagtctcatggtccctctcgacttattacatgagg 477

0v 736 ttattttatacagacaaataaggagctatctcagtaccttcacacaccacccatatcccc- 794

Db 478 ttctcttattatggc-acagggaatcagatccagcatcttcaaggacccaaccttcccca 536

795 -gagtcctcaccagtgagaaagagcctcccaggctatcttactactgagtgagtttttgccttc 853

db 537 aaagcatctccagaacaaggccacacagaccatcctgaatcctcatgaccttctttatc 596




```
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 138592
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138592
```

```
Query Match 2.9%; Score 32; DB 6; Length 738;
Best Local Similarity 100.0%; Pred. No. 5.1e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 723 ACAAGGTGATGAGGGCCTCTCCATCTGCACC 692
```

```
RESULT 3
US-10-027-632-325100/c
; Sequence 325100, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325100
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-325100
```

```
Query Match 2.9%; Score 32; DB 6; Length 738;
Best Local Similarity 100.0%; Pred. No. 5.1e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
```

```
Db 723 ACAAGGTGATGAGGGCCTCTCCATCTGCACC 692
|||||
```

```
RESULT 4
US-10-027-632-324938
; Sequence 324938, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 324938
; LENGTH: 916
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324938
```

```
Query Match 2.9%; Score 32; DB 6; Length 916;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 91 acaaggtgatgagggcctctccatctgcacc 122
```

```
RESULT 5
US-10-027-632-325047
; Sequence 325047, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325047
```

LENGTH: 916
TYPE: DNA
ORGANISM: Human
US-10-027-632-325047

Query Match 2.9%; Score 32; DB 6; Length 916;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 acaaggtgatgaggggcctctccatctgcacc 421
|||||
Db 91 acaaggtgatgaggggcctctccatctgcacc 122

RESULT 6
PCT-US02-09923-131
; Sequence 131, Application PC/TUS0209923
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael
; APPLICANT: YAO, Monique G.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: ISON, Craig H.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: BECHA, Shanya
; APPLICANT: LU, Yan
; APPLICANT: ARVIZU, Chandra P.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: WALSH, Roderick T.
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: GRUL, Richard C.
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PF-0933 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09923
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 60/280,683; 60/283,714; 60/287,266; 60/285,336
; PRIOR FILING DATE: 2001-03-30; 2001-04-13; 2001-04-27; 2001-04-20
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PERL Program
; SEQ ID NO 131
; LENGTH: 1097
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7476246CBI
PCT-US02-09923-131

Query Match 2.9%; Score 32; DB 1; Length 1097;
Best Local Similarity 100.0%; Pred. No. 5.3e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 acaaggtgatgaggggcctctccatctgcacc 421
|||||
Db 351 acaaggtgatgaggggcctctccatctgcacc 382

RESULT 7
PCT-US02-09188-1323
; Sequence 1323, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09188
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1323
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1323

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aaggtgatgaggggcctctccatctgcacc 421
|||||
Db 1 aaggtgatgaggggcctctccatctgcacc 30

RESULT 8
PCT-US02-09188-1326
; Sequence 1326, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09188
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1326

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aaggtgatgaggggcctctccatctgcacc 421
|||||
Db 1 aaggtgatgaggggcctctccatctgcacc 30

RESULT 9
PCT-US02-09188-1342
; Sequence 1342, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT

; CURRENT APPLICATION NUMBER: PCT/US02/09188
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1342
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1342

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 10
PCT-US02-09188-1343
; Sequence 1343, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09188
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1343
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1343

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 11
PCT-US02-09370-1381
; Sequence 1381, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1400
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens

; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1381
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1381

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 12
PCT-US02-09370-1384
; Sequence 1384, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1384
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1384

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 13
PCT-US02-09370-1400
; Sequence 1400, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1400
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens

PCT-US02-09370-1400

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 14

PCT-US02-09370-1401
; Sequence 1401, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1401
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1401

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 15

PCT-US02-09922-872
; Sequence 872, Application PC/TUS0209922
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS955PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09922
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1117
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 872
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09922-872

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

Search completed: July 21, 2002, 05:31:07
Job time: 5576 sec

